

Modelação de propriedades de pastas de betão autocompactável com comportamento heterocedástico

Diogo Filipe de Bastos Sousa Ribeiro

Orientado por:

Doutora Paula Manuela Lemos Pereira Milheiro de Oliveira
Professora Associada da FEUP

Doutora Sandra da Conceição Barbosa Nunes
Professora Auxiliar da FEUP

Tese apresentada para a obtenção do grau de
Mestre em Engenharia Matemática



Faculdade de Ciências
Universidade do Porto
Portugal
30 de Novembro de 2015

Para os meus pais Armando e Maria de Fátima

Agradecimentos

A realização desta dissertação só foi possível devido a todos aqueles que de uma forma ou de outra contribuíram para a realização da mesma, a todos os meus mais sinceros agradecimentos.

Agradeço desde logo e em primeiro lugar às minhas orientadoras Doutora Paula Manuela Lemos Pereira Milheiro de Oliveira e à Doutora Sandra da Conceição Barbosa Nunes pelo apoio e paciência revelada ao longo do período de realização desta tese.

Também gostaria de agradecer ao Doutor João Nunes Tavares e ao Doutor Carlos Miguel Menezes pela inspiração e ensinamentos e aqui deixar uma palavra de apreço ao Doutor Sílvio Marques de Almeida Gama pelo gosto que me incutiu na área de programação.

A todos os colegas que me apoiaram e acompanharam durante todas as etapas da minha formação superior, o seu apoio e amizade foram fundamentais para chegar até aqui. Em especial ao Eng. Paulo Almeida, director da Pietec SA.

Por fim, e como os últimos são sempre os primeiros, agradeço do fundo do coração, aos meus pais e irmã Solange, pelo apoio demonstrado ao longo destes anos, e minhas sobrinhas Lara, Sara e Sofia.

Resumo

O presente estudo foi integrado nas actividades da Unidade de Investigação CONSTRUCT, tendo sido desenvolvido no laboratório de Betão Estrutural da Faculdade de Engenharia da Universidade do Porto. O principal objectivo do trabalho consiste em contribuir para a compreensão da influência de variações na composição de pastas de cimento, da dispersão dos resultados obtidos com o ensaio do Cone de Marsh e a relação destes com a água livre avaliada através do ensaio de centrifugação a que se submeteram as mesmas pastas.

Para desenvolver este estudo foi elaborado um plano experimental tendo em vista a obtenção de uma amostra representativa de pastas de cimento com diversos rácios água/cimento e dosagens de superplastificante. As pastas de cimento, consistindo numa mistura de cimento, água, superplastificante e filler calcário foram sujeitas a um conjunto de três ensaios (ensaio do cone de Marsh, ensaio de centrifugação e ensaio de espalhamento, este último executado para controlo da qualidade da pasta) permitindo a observação de variáveis aleatórias entre as quais "tempo de escoamento" e "quantidade de água livre".

Com vista a estabelecer uma relação estatística entre o tempo de escoamento obtido no ensaio do cone de Marsh e a água livre obtida pelo ensaio de centrifugação foi efectuada uma análise estatística recorrendo ao modelo de regressão gaussiano, usando o método dos mínimos quadrados ponderados, para solucionar o problema de heterocedasticidade presente no tempo de escoamento quando vista como dependente da água livre.

Analisando os resultados experimentais obtidos concluiu-se que o valor do tempo de escoamento das caldas diminui com o aumento da razão água/cimento utilizada na sua preparação e foi estabelecido um modelo estatístico entre o tempo de escoamento de uma pasta de cimento no ensaio do cone de Marsh com a água livre obtida pelo ensaio de centrifugação.

Palavras-Chave: Água livre; Cone de Marsh; Heterocedasticidade; Modelo linear gaussiano; Método dos mínimos quadrados ponderado; Pastas de cimento; Tempo de escoamento.

Abstract

The current study was developed within CONSTRUCT Research Unit, and carried out in the Structural Concrete Laboratory of the Faculty of Engineering of the Porto University. The aim of this study is to understand the influence of cement paste mix-proportions on the results Marsh Cone test and its relationship with the free water content measured by the centrifuge test.

An experimental plan was carried out to characterize cement pastes with varying water/cement ratios and superplasticizer dosages. The cement pastes, consisting of a mixture of cement, water, superplasticizer and limestone filler were subjected to a set of three tests (Marsh cone test, centrifugation test and slump test, the latter was carried out for the paste quality control) allowing the observation of random variables including "flow time" and "amount of free water."

To establish a statistical relationship between flow time obtained in the Marsh cone test and free water obtained by the centrifugation test a statistical analysis using the Gaussian regression model was performed applying the method of weighted least squares to solve the problem of heteroscedasticity present in the flow time when seen as dependent on the free water.

Analyzing the obtained experimental results it was concluded that the flow time decreases with increasing water/cement ratio and a statistical model was established between the flow time of a cement paste in the Marsh cone and the free water obtained by the centrifugation test.

Keywords: Cement pastes; Free water; Flow time; Gaussian linear model; Heteroscedasticity; Marsh cone; Method of weighted least squares;

Lista de Figuras

3.1	Procedimentos para a determinação do ponto de saturação pelo método de AFREM	37
3.2	Procedimentos para a determinação do ponto de saturação pelo método de AFREM	37
3.3	Procedimentos para a determinação do ponto de saturação pelo método de AFREM	38
3.4	Procedimentos para a determinação do ponto de saturação pelo método de AFREM	38
3.5	Procedimentos para a determinação do ponto de saturação pelo método de AFREM	39
3.6	Rácio $w/p=0.42$	39
3.7	Procedimentos para a determinação do ponto de saturação pelo método de AFREM	39
4.1	Gráfico de dispersão dos resíduos estudentizados contra a variável independente e contra os valores ajustados do modelo	43
4.2	Gráfico dos resíduos estudentizados para aferir da forma da heterocedasticidade, onde os nomes dos eixos são respectivamente na horizontal a variável independente $z = x$ e $z = y\hat{\beta}$, e na vertical os valores dos resíduos estudentizados.	66
4.3	Gráfico dos resíduos estudentizados para aferir da forma da heterocedasticidade, onde os eixos são respectivamente na horizontal a variável independente $z = x$ e $z = y\hat{\beta}$, e na vertical os valores dos resíduos estudentizados quadráticos.	67
4.4	Gráfico parcial de resíduos do modelo 4.43.	68
4.5	(a)Gráfico de dispersão em que a variância da resposta é proporcional ao quadrado da média, (b) Gráfico de dispersão dos mesmos dados depois de aplicada a transformação logarítmica	68
4.6	Transformação de Box-Cox	69
4.7	Regra de Mosteller e Tukey	69
4.8	Diagrama de dispersão do valor-p do teste de Breusch-Pagan consoante o tamanho da amostra	70
4.9	Diagrama de dispersão da taxa de erro para o teste de Breusch-Pagan consoante o tamanho da amostra	70
4.10	Gráfico de dispersão entre X e Y para o modelo (4.79)	74
4.11	Gráfico dos resíduos para o modelo (4.79)	75
4.12	Gráfico de dispersão entre x e y para o modelo (4.91)	77

4.13	Gráfico dos resíduos para o modelo (4.91)	78
4.14	Gráfico de dispersão entre x e y para o modelo 4.98	79
4.15	Gráfico dos resíduos para o modelo 4.98	80
4.16	Gráfico de dispersão entre x e y para o modelo 4.101	82
4.17	Gráfico dos resíduos para o modelo 4.101	82
5.1	Gráfico de dispersão dos dados	89
5.2	Gráfico de dispersão dos dados e a recta de regressão definida pelo modelo (5.2)	91
5.3	Histograma dos resíduos do modelo (5.2)	92
5.4	Histograma dos resíduos do modelo (5.4)	94
5.5	Gráfico QQ dos resíduos do modelo (5.4)	95
5.6	Gráfico de dispersão dos dados e recta de regressão definida pelo modelo (5.4)	96
5.7	Gráfico de Componente+Resíduos do modelo (5.4)	96
5.8	Gráfico dos resíduos de Pearson do modelo (5.4)	97
5.9	Gráfico de dispersão dos dados com a curva de regressão definida pelo modelo (5.6), com os respectivos intervalos de confiança para os valores ajustados da regressão e para os valores previstos.	98
5.10	Gráfico dos resíduos de Pearson do modelo (5.6)	99
5.11	Gráfico dos resíduos estudentizados para aferir da forma da heterocedasticidade, onde os eixos são respectivamente na horizontal a variável independente $Wfree$ e na vertical o valor absoluto dos resíduos estudentizados	100
5.12	Gráfico dos resíduos estudentizados para aferir da forma da heterocedasticidade, onde os eixos são respectivamente na horizontal a variável independente $Wfree$ e na vertical o quadrados dos resíduos estudentizados	100
5.13	Gráfico das variáveis do modelo (5.11) contra os resíduos de Pearson.	102
5.14	Gráfico de dispersão dos dados com a curva de regressão definida pelo modelo (5.11), com os respectivos intervalos de confiança para os valores ajustados da regressão e para os valores previstos.	103
5.15	Gráfico QQ normais para os resíduos do modelo (5.11)	105
5.16	Gráfico dos Resíduos Estudentizados com os intervalos de confiança para o modelo (5.11)	105
5.17	Gráficos de diagnóstico para as estatísticas do modelo (5.11)	106
5.18	Gráfico QQ normais para os resíduos do modelo (5.12)	107
5.19	Gráfico dos Resíduos Estudentizados com os intervalos de confiança para o modelo (5.12)	107
5.20	Gráficos de diagnóstico para as estatísticas do modelo (5.12)	109
5.21	Gráfico dos resíduos de Pearson do modelo (5.13)	111
5.22	Gráfico Componente+Resíduo do modelo (5.13)	112
5.23	Gráfico de dispersão dos dados e curva de regressão definida pelo modelo (5.13), com os respectivos intervalos de confiança para os valores ajustados da regressão e para os valores previstos.	113
5.24	Gráfico da transformação de Box-Cox para a variável resposta do modelo (5.15)	114
5.25	Gráfico QQ normais para o modelo (5.17)	116
5.26	Gráfico dos resíduos estudentizados com os intervalos de confiança modelo (5.17)	117
5.27	Gráfico das variáveis do modelo (5.17) contra os resíduos de Pearson.	120

5.28	Gráfico de dispersão dos dados com a curva de regressão definida pelo modelo (5.17), com os respectivos intervalos de confiança para os valores ajustados da regressão e para os valores previstos.	120
5.29	Gráfico de dispersão dos dados com a curva de regressão definida pelo modelo 5.13, com os respectivos intervalos de confiança para os coeficientes de regressão e para os valores previstos em comparação com uma modelação da variância por métodos não paramétricos	121
5.30	Gráfico de dispersão dos dados com a curva de regressão definida pelo modelo (5.13), com os respectivos intervalos de confiança para os coeficientes de regressão e para os valores previstos em comparação com uma modelação da variância por métodos não paramétricos	121
6.1	Gráfico de dispersão dos dados do estudo efectuado em Nunes et. al [110] e curva de regressão definida pelo modelo (5.13), com os respectivos intervalos de confiança para os coeficientes de regressão e para os valores previstos. . .	125

Lista de Tabelas

2.1	Tipos de cimentos (adaptado da norma NP EN 197-1).	23
2.2	Constituição do clínquer do cimento Portland	24
2.3	Adições para o betão (Adaptado de European Guidelines, 2005)	25
3.1	Propriedades químicas e físicas do cimento e filler calcário	32
3.2	Características segundo informação do fornecedor do superplastificante. . . .	33
5.1	Sumário dos dados	88
5.2	Sumário dos dados actualizada	88
5.3	Matriz de Correlação de Pearson	88
5.4	Matriz de Correlação de Spearman	88
5.5	Resultados obtidos para o modelo de regressão linear (5.1)	90
5.6	Resultados do modelo de regressão linear (5.3), coluna(2), e comparação com o modelo (5.1), coluna (1)	93
5.7	Tabela do teste de curvatura para os resíduos do modelo (5.4)	93
5.8	Resultados do modelo de regressão linear (5.5), coluna(2) e comparação com o modelo (5.3), coluna (1)	97
5.9	Tabela do teste de curvatura para os resíduos do modelo (5.6)	98
5.10	Testes de homocedasticidade para o modelo (5.6)	99
5.11	Resultados do modelo de regressão linear (5.7), coluna (1), e (5.8), coluna (2)	101
5.12	Resultados do modelo de regressão linear (5.5) usando o modelo de variância (5.9), coluna(2), e comparação com o modelo (5.5), coluna (1), e (5.5) usando o modelo de variância (5.10), coluna(3)	101
5.13	Tabela do teste de curvatura para os resíduos do modelo (5.11)	102
5.14	Testes de normalidade dos resíduos	104
5.15	Teste de autocorrelação dos resíduos	104
5.16	Testes aos Outliers	108
5.17	Valores das variáveis excluídas na construção do modelo (5.12)	108
5.18	Resultados do modelo de regressão linear (5.12) retirados os outliers, coluna(2), e comparação com o modelo (5.11), coluna (1)	108
5.19	Resultados do modelo de regressão linear (5.13) uma vez retirados os pontos de influência, coluna(2), e comparação com o modelo (5.12), coluna (1) . . .	110
5.20	Tabela do teste de curvatura para os resíduos do modelo (5.12)	111
5.21	Resultados do modelo de regressão linear (5.13), coluna(1), e comparação com o modelo (5.17), coluna (2)	115
5.22	Resultados do modelo de regressão linear (5.16)	115

5.23	Resultados do modelo de regressão linear (5.13), coluna(1), em comparação com a variância modelada por: regressão Quantil (2), regressão Loess (3), regressão por núcleos (4), regressão Lasso (5).	118
5.24	Resultados do modelo de regressão linear usando estimadores HCCM: HC0(1), HC1 (2), HC2 (3), HC3 (4), HC4 (5).	119
A.1	Tabela de registro dos valores dos ensaios	127

Conteúdo

Agradecimentos	3
Resumo	5
Abstract	7
Lista de Figuras	9
Lista de Tabelas	13
1 Introdução	17
2 O material Betão	21
2.1 Betão Auto-Compactável	21
2.2 Cimento	22
2.3 Adições	24
2.4 Adjuvantes	26
3 Procedimento experimental	31
3.1 Preparação da amostra	31
3.2 Proporções das misturas	33
3.3 Caracterização dos materiais utilizados	34
3.4 Descrição dos ensaios adoptados	34
3.5 Controlo dos resultados	36
4 Heterocedasticidade dos resíduos da regressão	41
4.1 Consequências de ignorar a heterocedasticidade	44
4.2 Método dos mínimos quadrados generalizados	46
4.3 Método dos mínimos quadrados ponderados	47
4.4 Resíduos	51
4.5 Testes à forma funcional da regressão	54
4.6 Transformações para a homocedasticidade	56
4.7 Pré-Testes	58
4.7.1 Teste de White	58
4.7.2 Teste do coeficiente de correlação de Spearman	59
4.7.3 Teste de Goldfeld-Quandt	60
4.7.4 Teste de Breusch-Pagan	62

4.7.5	Teste de Szroeter	62
4.7.6	Teste de Monte Carlo	64
4.8	Modelação da heterocedasticidade	71
4.8.1	Variância constante entre subgrupos	71
4.8.2	Desvio padrão de Y é uma função linear de uma ou mais variáveis independentes	72
4.8.3	Variância de Y é uma função linear de uma ou mais variáveis independentes	75
4.8.4	Variância de Y é proporcional a uma potência do valor esperado . . .	76
4.8.5	Heterocedasticidade multiplicativa	78
4.8.6	Estimadores de White	83
4.9	Comentários	84
5	Aplicação na modelação da fluidez de pastas de betão	87
5.1	Breve discussão dos dados	87
5.2	Uma primeira abordagem de modelação	89
5.3	Modelo log-polinomial	92
5.4	Modelo hiperbólico	113
5.5	Discussão dos resultados	116
6	Conclusões e comentários finais	123
A	Tabela de Registro dos Resultados dos Ensaio	127
	Bibliografia	129

Capítulo 1

Introdução

O sector da construção tem um papel preponderante no desenvolvimento da sociedade, visto ser o responsável pela construção das infraestruturas necessárias ao desenvolvimento da mesma. Este contributo, no entanto, acarreta aspectos negativos, pois o sector da construção, nomeadamente, a indústria do cimento e betão é responsável pelo consumo de matérias primas não renováveis e pelo consumo de energia e consequente emissão de gases de estufa. Dada a evolução do conhecimento, e em virtude de uma sociedade preocupada com o meio ambiente, a segurança da obra e os aspectos económicos, é fundamental investir na fase de concepção, construção, protecção e cura do betão, pois, tal como proposto por Sitter e designado por "Lei dos Cinco" (ver [65]):

1 Euro investido na fase A, corresponde a 5 Euros na fase B, 25 na fase C e a 125 na fase D.

em que cada uma das fases consiste em:

- A.** Projecto, construção e cura do betão;
- B.** Processos de iniciação da corrosão em curso, mas com as fases de propagação e portanto de deterioração ainda não iniciadas;
- C.** Propagação da corrosão iniciada;
- D.** Estado de propagação avançada.

O betão é um dos materiais de construção mais usados em engenharia civil, a par do aço. No entanto, o engenheiro tem um conhecimento mais profundo do aço do que do betão, pois o aço é produzido sob condições muito controladas e com propriedades certificadas pelo fabricante de acordo com as normas vigentes. Assim sendo o engenheiro ou projectista apenas necessita de especificar a classe do aço de acordo com o que as normas ditam em função das características individuais de cada obra, sendo apenas obrigado a verificar na obra a correcta e ligações das armaduras. No caso do betão, o mesmo já não se verifica, pois, apesar do cimento ter normas específicas de fabrico e controlo de qualidade, este não é o material de construção mas sim o betão, e a qualidade do betão depende sobremaneira da mão de obra que o produz, o coloca e o cura. Portanto o engenheiro deve ter em consideração que esta condicionante pode facilmente prejudicar as propriedades do betão fabricado em

obra e desviar-se consideravelmente do pretendido no projecto. Outro aspecto a ter em consideração é o meio onde a obra está inserida, pois o betão interage com o ambiente que o rodeia e o seu comportamento ao longo da sua vida útil é influenciado pelo ambiente. Daqui se conclui que, para obter um betão de qualidade, é necessário conhecer as propriedades de cada um dos seus constituintes, as interacções entre cada um desses constituintes e as interacções com o meio ambiente e controlar as fases de produção, transporte, colocação e compactação, isto é, o seu comportamento em estado fresco, assim como endurecido. Só assim se pode garantir que o betão terá as características necessárias à obra em questão. É devido a estas particularidades que surge o problema que origina esta tese.

O presente trabalho tem como objectivo a modelação matemática de uma relação entre variáveis que são medidas através de dois ensaios laboratoriais usados em engenharia civil para avaliar das características de pastas de betão em estado fresco, nomeadamente, o ensaio do cone de Marsh e o ensaio de centrifugação, são elas o tempo de escoamento e a quantidade de água livre por m³. A suspeita da existência de uma relação entre estes dois ensaios surge pela primeira vez com o estudo efectuado por Nunes et. al [110], e a questão foi deixada em aberto por esses autores, uma vez que as técnicas de regressão clássica falharam no estabelecimento de um modelo estatístico que garantisse previsões de qualidade, fruto da heterocedasticidade dos resíduos do modelo então encontrado. O trabalho de Nunes et. al [110] deixou claro que se impunha uma investigação cuidada pela via dos modelos de regressão com resíduos heterocedásticos, começando pela aplicação de testes à presença de heterocedasticidade e à modelação da variância. Este assunto é tradicionalmente tratado na área de economia, mas menos desenvolvido nas áreas das engenharias.

Torna-se então necessário definir heterocedasticidade. Esta consiste na violação da hipótese de variância constante dos resíduos do modelo clássico de regressão linear, dado que este pressupõe que a variância de cada resíduo, condicional aos valores das variáveis explicativas, é constante, ou seja, a relação é homocedástica. Por outras palavras, a heterocedasticidade apresenta um padrão não constante da distribuição dos resíduos do modelo em torno da recta $y = 0$.

White [153] é o artigo mais citado, embora na área de economia, no período compreendido entre 1970 e 2005 (ver [91]) e inaugurou uma nova era na inferência estatística para modelos heterocedásticos. Este interesse sobre a heterocedasticidade tem sido renovado com o desenvolvimento dos modelos ARCH e GARCH (ver [58]), com a construção de testes estatísticos com base em técnicas de Monte Carlo que evitem erros de Tipo I (ver [58, 71]), com propostas de modelação da heterocedasticidade por redes neurais artificiais (ver [34]) ou métodos não paramétricos (ver [95]). No presente trabalho iremos centrar-nos em modelos heterocedásticos mais clássicos, tais como os descritos por Harvey [83], Glejser [68], Carrol [39], entre muitos outros [58].

Um obstáculo encontrado na elaboração desta tese foi a inexistência de bibliografia sobre estudos estatísticos que relacionassem o ensaio do cone de Marsh e o ensaio de centrifugação. Que seja do nosso conhecimento, só existe o estudo anteriormente referido [110]. No que se refere ao ensaio do cone de Marsh e a propriedades de pastas de cimento, existem os mais diversos estudos, tais como [2, 50, 62, 121, 122, 109], que tratando cada um deles aspectos diferentes do comportamento dos betões e dos ensaios para observar esses comportamentos.

Para a resolução do nosso problema estabeleceu-se então um plano de trabalhos que contemplou:

- a replicação do procedimento experimental descrito em Nunes et. al (ver [110]) adaptado à pesquisa das propriedades do betão que nos interessou;
- a pesquisa de referências bibliográficas que nos pudessem guiar no que diz respeito, tanto a modelos de regressão heterocedásticos, como a engenharia de materiais de construção no que respeita ao comportamento do cimento e betão;
- terminando com a proposta de um modelo estatístico para a relação que se pretendia estudar.

Esta tese está organizada em 6 capítulos:

1. este capítulo que constitui uma introdução;
2. o Capítulo 2 em que apresentamos brevemente o material com que vamos trabalhar no estudo, o betão;
3. o Capítulo 3 em que é feita a descrição do programa experimental, onde se apresentam todos os procedimentos experimentais realizados na produção e caracterização das pastas para betão auto-compactável;
4. no Capítulo 4 é feita uma breve descrição do modelo linear gaussiano, das consequências de ignorar a heterocedasticidade, a descrição de testes para avaliar a presença de dados heterocedásticos e de formas
5. no Capítulo 5 é feita a apresentação e discussão dos resultados obtidos, confrontando esses resultados com a revisão bibliográfica efectuada;
6. no Capítulo 6, o último, apresentamos as conclusões, tecemos alguns comentários finais e deixamos sugestões para trabalhos futuros.

Capítulo 2

O material Betão

Neste capítulo será feito um resumo da produção do cimento, os seus custos ambientais, a caracterização dos diversos tipos de cimento e materiais adicionados ao mesmo, como também a descrição do que são superplastificantes, suas propriedades e o motivo porque são usados na construção civil.

2.1 Betão Auto-Compactável

O betão auto-compactável surgiu no Japão no início da década de 80 do século passado devido ao trabalho de investigação do Professor Okamura da Universidade de Tóquio em 1986 (ver [74]). A motivação consistiu no desenvolvimento de um betão auto-compactável que dispensasse a intervenção humana na fase de colocação, levando à eliminação da necessidade de mão de obra qualificada e, por sua vez, a uma melhoria da uniformidade e homogeneidade do betão aplicado nas estruturas. Os subseqüentes desenvolvimentos desta tecnologia permitiram posteriormente a sua aplicação na construção de pontes e estruturas de edifícios. Ou seja, a aplicação de um betão que conseguia preencher completamente os moldes, passando por zonas densamente armadas sem a ocorrência do bloqueio dos agregados nem segregação, unicamente pelo efeito do próprio peso e sem recurso a qualquer método de compactação.

Tal como é descrito em [33], a aplicação do betão auto-compactável, tem aspectos positivos e negativos. Como vantagens do betão auto-compactável, podem enunciar-se:

- Redução do custo de produção pois, visto não requerer a utilização de vibradores, não é necessário mão-de-obra especializada, para a realização da compactação e de modo a assegurar o total preenchimentos das cofragens e envolvimento das armaduras;
- redução do período de construção, visto não ser necessário realizar a compactação do betão, o processo de betonagem torna-se mais rápido;
- melhoria das condições de trabalho e do meio ambiente: a eliminação da vibração no processo de colocação do betão, resulta numa diminuição do ruído associado;
- maior liberdade na concepção estrutural passa a ser possível a utilização de elementos estruturais como as mais diferentes formas, mesmo que complexas;

- maior durabilidade e melhor acabamento pois, ao contrário do betão corrente, ao não necessitar de vibração gera estruturas homogéneas e como tal com um maior período de vida útil e menor necessidade de manutenção.

Quanto aos inconvenientes do betão auto-compactável podem referir-se aos seguintes:

- Elevados custos: ao contrário do betão convencional, o betão auto-compactável possui na sua composição uma elevada quantidade de finos, que por sua vez, é o componente mais dispendiosa no custo total do betão;
- maior sensibilidade a variações dos constituintes: o betão auto-compactável é bastante influenciado pela variabilidade dos seus materiais constituintes;
- cofragens estanques: como se trata de um betão mais fluido, é necessário ter cuidados especiais com as cofragens, quer ao nível da estanquidade, quer ao nível da sua solidez e limpeza.

No início a sua utilização era restringida a projectos onde a colocação do betão era de difícil execução devido a grande densidade de armaduras. Com a construção dos primeiros edifícios, pilares e vãos de pontes, tornou-se claro que o leque de aplicações do betão auto-compactável é bem mais vasto, o que fez aumentar gradualmente a sua utilização. Em seguida são descritas as principais aplicações práticas do betão auto-compactável,[111]:

- Fundações, pilares, lajes, vigas e peças delgadas;
- peças com alta densidade de armadura;
- túneis (arcos de sustentação) e Pontes (maciços de ancoragem, pilares, vigas);
- peças com dimensões reduzidas e estruturas pré-moldadas;
- reabilitação e/ou reforço de estrutura.

2.2 Cimento

Cimento Portland é a denominação convencionada mundialmente para o material usualmente conhecido na construção civil como cimento. Segundo a norma EN 191-1 o cimento é um ligante hidráulico, isto é, um material inorgânico, finamente moído, que convenientemente amassado com água, forma uma pasta que devido a reacções de hidratação, faz presa, endurece e permanece mecanicamente resistente e estável, tanto ao ar como na água, em condições normais de aplicação. Um cimento que esteja de acordo com a norma europeia tem a designação de CEM.

Por definição, é um aglomerante hidráulico resultante da mistura homogénea de clínquer Portland, gesso e adições finamente moídas. Aglomerante porque tem a propriedade de unir outros materiais. Hidráulico porque hidrata ao se misturar com água e depois de endurecido ganha características de rocha artificial, mantendo suas propriedades.

As matérias primas utilizadas na fabricação de cimento devem conter Cálcio, Silício, Alumínio e Ferro, pois são estes os elementos químicos que, combinados, vão produzir compostos hidráulicos activos. Dependendo dos materiais adicionados ao clínquer o cimento Portland tem diversas classificações descritas na tabela 2.1. As adições são outras matérias-primas que, misturadas ao clínquer na fase de moagem, permitem a fabricação dos diversos tipos de cimento Portland hoje disponíveis no mercado. Essas outras matérias-primas são o gesso, as escórias de alto-forno, os materiais pozolânicos e os materiais carbonatos.

O clínquer é o principal constituinte na composição de cimentos Portland, tem como matérias-primas o calcário e a argila e é a fonte de silicato tricálcico e silicato dicálcico. Estes compostos são os que conferem a característica de ligante hidráulico e estão directamente relacionados com a resistência mecânica do material após a hidratação.

Tipos de cimentos principais	Constituintes
CEM I – Cimento Portland	Clínquer 95-100%
CEM II – Cimento Portland Composto	Clínquer > 65%
CEM III – Cimento de Alto Forno	Clínquer + 35-95% de escórias de alto forno
CEM IV – Cimento Pozolânico	Clínquer 45% a 89% + 11% a 55% de sílica de fumo, pozolana ou cinzas volantes
CEM V – Cimento Composto	Clinquer > 20% + escória >18% + pozolana e/ou cinzas volantes siliciosas

Tabela 2.1: Tipos de cimentos (adaptado da norma NP EN 197-1).

Processo de fabricação

Embora seja um dos mais antigos materiais de construção, o processo de produção do cimento é uma combinação de fórmulas tradicionais e alta tecnologia em equipamentos. Em geral, uma fábrica de cimento situa-se nas proximidades de uma formação calcária, visto que 70 a 80% da matéria prima é calcário.

A fabricação do cimento Portland baseia-se nas seguintes etapas:

- Preparação do cru:
 1. extracção do calcário;
 2. britagem;
 3. pré-homogeneização, que consiste em formar uma pilha de armazenamento com argila e remoção por cortes verticais com 75% de brita calcária e 25% de argila;
 4. moagem e homogeneização;
 5. correcção com aditivos.
- Cozedura:
 1. armazenamento nos silos alimentadores do forno;
 2. pré-aquecimento;
 3. entrada no forno rotativo.
- Arrefecimento, adição de gesso e moagem;

- Ensilagem e Ensacagem;

As matérias-primas necessárias para a produção de cimento, carbonato de cálcio, sílica, alumínio e minério de ferro, são geralmente extraídas de rocha calcária ou argila. Em seguida, são trituradas e transportadas para a fábrica onde são armazenadas e homogeneizadas.

A moagem das matérias-primas produz um pó fino conhecido como cru que é pré-aquecido e em seguida introduzido num forno rotativo. O material é aquecido a uma temperatura de 1.500°C através de uma chama a 2.000°C , antes de ser subitamente arrefecido por injeção de ar. Assim é produzido o clínquer, o material básico necessário para a produção de todos os tipos de cimento.

Uma pequena quantidade de gesso, 3 a 5%, é adicionada ao clínquer para regular o processo de endurecimento do cimento. A mistura é então finamente moída para se obter o cimento puro. Durante esta fase, diferentes materiais minerais, chamados de "adições", podem ser adicionados além do gesso. Usadas em variadas proporções, essas adições, que podem ser recursos naturais ou sub-produtos industriais, dão ao cimento propriedades específicas como redução de impermeabilidade, resistência a sulfatos e ambientes agressivos, melhor desempenho e acabamento.

Finalmente, o processo produtivo do cimento termina e o produto final é armazenado em silos, antes de ser enviado a granel ou em sacos para consumo.

O clínquer de cimento Portland é constituído por:

Constituinte	Percentagem
Óxido de cálcio (CaO)	60 a 70%
Sílica (SiO_2)	20 a 25%
Alumina (Al_2O_3)	2 a 9%
Óxido de ferro III (Fe_2O_3)	1 a 6%
Óxido de magnésio (MgO)	0 a 2%

Tabela 2.2: Constituição do clínquer do cimento Portland

2.3 Adições

As adições são materiais inorgânicos finos, moídos ou não, de granulometria menor ou igual à do cimento, que podem ser utilizados no betão com a finalidade de melhorar as suas propriedades, ou mesmo para adquirir propriedades especiais, tais como, a resistência à compressão através da redução do calor de hidratação, a trabalhabilidade e a durabilidade. Para além destes factores, as adições contribuem para a sustentabilidade e redução dos impactos ambientais associados à indústria do betão através da reutilização de subprodutos industriais em substituição parcial do cimento, reduzindo assim a energia e poluição associado à produção de betão e a reutilização de materiais que de outra forma teriam como destino final o aterro.

O documento normativo NP EN 206-1 (2005), referente às especificações técnicas do betão, refere que as adições minerais podem ser classificadas em dois tipos, de acordo com a sua interacção com a água:

Tipo	Adição	Exemplo
I	Adições quase inertes	Fíleres minerais conforme NP EN 12620:2004 Pigmentos conforme EN 12878.
II	Adições pozolânicas	Cinzas volantes conforme NP EN 450:1995 Sílica de fumo conforme LNEC E – 377:1993.
	Hidráulicas latentes	Escória granulada de alto forno conforme LNEC E – 375:1993.

Tabela 2.3: Adições para o betão (Adaptado de European Guidelines, 2005)

Em alternativa ao estabelecido na NP EN 206-1 para efeitos de utilização de adições, a especificação E 464 introduz uma outra metodologia para a utilização das adições como ligantes, trata-se do conceito de mistura, o qual tem as seguintes condições de utilização:

1. o cimento a utilizar seja do tipo CEM I ou CEM II/A e de classe de resistência 42,5 ou superior;
2. as adições sejam do tipo I de origem calcária (satisfazendo a especificação E 466) ou do tipo II;
3. a composição da mistura satisfaça os limites de composição de um dos cimentos previstos no Quadro 1 da NP EN 197-1;
4. a proporção da sílica de fumo não ultrapasse 11% da massa de clínquer na mistura.

Quando for utilizado este conceito, os termos “dosagem de cimento” e “razão água/cimento” devem ser substituídos pelos termos “dosagem de ligante” e “razão água/ligante”, respectivamente.

Esta dissertação irá centrar-se só e apenas nas adições quase inertes (tipo I), nomeadamente, os filers calcários, sendo as suas propriedades descritas de seguida.

Filer calcário

Como mencionado acima, esta é uma adição do tipo I, quase inerte. A adição de filer calcário finamente moído é efectuada para diminuir a percentagem de vazios, porque os grãos ou partículas desses materiais têm dimensões adequadas para se alojar entre os grãos ou partículas dos demais componentes do cimento. É um material proveniente de rochas calcárias, cujo principal componente é o carbonato de cálcio. Este tipo de adição é bastante usada no betão auto-compactável, podendo dar excelentes propriedades reológicas e um bom acabamento superficial do betão endurecido. O ideal é que o tamanho das partículas de filer calcário seja inferior a 0,125 mm (ver [51]). Os filers calcários são especialmente adequados para o betão auto-compactável por garantirem uma maior uniformidade da granulometria da mistura, controlando melhor a quantidade de água na mistura conduzindo a melhoria da trabalhabilidade e diminuindo a exsudação. Tal como é afirmado em [111], nos últimos anos existe um aumento do uso de filer calcário no betão auto-compactável para estruturas com grandes massas de betão, onde ocorre frequentemente fendilhação devido a dilatação térmica. A utilização de filer calcário reduz o calor de hidratação diminuindo o aumento de temperatura do betão.

Adições pozolânicas e hidráulicas latentes

Um material com propriedades pozolânicas ou hidráulicas latentes é um material que permite a formação de silicatos e aluminatos de cálcio hidratados, isto é, são verdadeiros cimentos, mas reagem com lentidão quando usados isoladamente, portanto não têm aplicação prática. A activação é feita, quando se junta soda, cal, silicato de sódio ou cimento Portland, pois o pH sobe para um valor de 12, activando as reacções de hidratação das escorias. A hidratação do cimento produz uma grande quantidade de hidróxido de cálcio que é capaz de activar a reacção deste tipo de adições.

As escorias de alto forno são um subproduto do fabrico de ferro fundido, constituído por cal, sílica e alumina, que se encontram no estado vítreo.. É um material que é obtido pela combinação de ganga dos minérios dos metais com fundentes e cinzas de carvão, que quando arrefecidos bruscamente e finamente moídos adquirem propriedades ligantes.

Pozolanas são produtos naturais ou artificiais constituídos por sílica e alumina. Apesar de não terem propriedades aglomerantes e hidráulicas (por si só), contêm constituintes que se combinam na presença de água e a temperaturas normais, com o hidróxido de cálcio para formarem produtos resistentes como os C-S-H (silicatos hidratados de cálcio). As vantagens de usar o cimento pozolânico são que estes têm baixo calor de hidratação e permite refinar a estrutura porosa por consumo do hidróxido de cálcio proveniente da hidratação do clínquer.

Cinzas volantes é um resíduo da queima do carvão que é arrastado nos gases de combustão das centrais térmicas. Em Portugal é obtido nas centrais térmicas de Sines e Pego. Existem dois tipos de cinzas volantes, a Cinza volante siliciosa que é constituída essencialmente por sílica e alumina com baixo teor de oxido de calcário reactivo ($<10\%$). E a Cinza volante calcária que é constituída por sílica, alumina e com um elevado teor de oxido de calcário reactivo($>10\%$).

Silica de fumo é um subproduto da preparação do silício ou de ligas de silício, especialmente ferro-silício. Provem da redução do quartzo de elevada pureza com carvão em fornos de arco eléctrico. Não deve ser usada em percentagem elevada em relação à quantidade de cimento pois pode produzir microfissuração a longo prazo.

2.4 Adjuvantes

São materiais de natureza mineral ou orgânica adicionados ao betão, em percentagem inferior a 5% da massa do cimento, NP EN 934-2 (2009), durante o processo de amassadura, para modificar as propriedades do betão fresco ou endurecido, no sentido do cumprimento dos requisitos específicos de cada obra como, por exemplo, o aumento da resistência, redução da permeabilidade, manutenção da consistência e redução da razão w/c .

Nos nossos dias, em virtude do custo do cimento, o emprego de adjuvantes é mesmo uma necessidade, apesar de não terem sempre um custo baixo, não representam necessariamente um custo adicional, pois da sua utilização podem resultar poupanças económicas, por exemplo, no custo de mão-de-obra para efectuar a compactação, no teor de cimento que seria necessário, ou na melhoria da durabilidade sem a utilização de medidas adicionais.

O desempenho de um determinado betão pode ser substancialmente melhorado com a utilização de adjuvantes, desde que sejam cumpridas as recomendações do fabricante e que a sua utilização tenha sido considerada na formulação da composição do betão.

Principais características, que se pretendem alcançar, com o uso de adjuvantes são:

- Melhorar a trabalhabilidade;
- acelerar a presa;
- retardar a presa;
- acelerar o endurecimento nas primeiras idades;
- aumentar a resistência aos ciclos de gelo/degelo;
- diminuir a permeabilidade aos líquidos;
- impedir a segregação e a sedimentação do cimento nas caldas de injeção;
- criar uma ligeira expansão no cimento, para ancoragem do pré-esforço;
- aumentar a aderência ao agregado e aos betões endurecidos;
- produzir betão colorido;
- produzir betão celular (leve);
- produzir propriedades fungicidas, germicidas e insecticidas;
- ajudar a bombear betão pobres;
- inibir a corrosão das armaduras.

A classificação dos adjuvantes é geralmente efectuada com base na sua actuação sobre as propriedades tecnológicas do betão. Esta classificação é muitas vezes difícil, pois muitos deles possuem simultaneamente várias características, pelo que se torna necessário determinar ou atribuir-lhe uma função principal, que permita a sua classificação, como:

- plastificantes e superplastificantes (redutores de água);
- aceleradores e retardadores de presa;
- aceleradores de endurecimento;
- hidrófugos (ou redutores da capilaridade);
- introdutores de ar, etc

Há diversas características a que os adjuvantes devem obedecer, nomeadamente a identificação, tal como a homogeneidade, cor, teor de sólido, entre outros, a compatibilidade com os cimentos e o comportamento no betão. Uma descrição mais pormenorizada destas características encontra-se na norma NP EN 934-1 (2008). Do conjunto de adjuvantes cobertos por esta norma, o presente trabalho versará sobre os superplastificantes.

Superplastificantes

Os superplastificantes são adjuvantes de nova geração, que podem ser adicionados às misturas de betão para fazê-las mais trabalháveis. De acordo com a norma NP EN 934-2 (2009), os superplastificantes são adjuvantes que, sem modificar a sua consistência, permitem reduzir fortemente a dosagem de água dum dado betão, ou que, sem modificar a dosagem de água aumentam consideravelmente o abaixamento ou o espalhamento, ou que produzem os dois efeitos simultaneamente. Estes são vastamente usados em diferentes campos da indústria para melhorar as propriedades reológicas de partículas suspensas, especialmente em aplicações de cimento. A sua adição permite a redução da razão água/cimento, melhorando a trabalhabilidade de misturas frescas e o desempenho de pastas, argamassas e betão (ver [155]). Apesar da sua utilização generalizada, estes polímeros continuam a ser objecto de estudo, pois existem muitas falhas no completo entendimento do seu desempenho (ver [72]).

O efeito dos superplastificantes é determinado, fundamentalmente, por três fenómenos de natureza físico-química, dispersão, adsorção, a adesão de moléculas de um fluido a uma superfície sólida, e potencial de repulsão intermolecular zeta, que resultam na desfloculação e dispersão das partículas de cimento, conferindo-lhes uma forte carga eléctrica negativa de modo a repeli-las umas das outras.

Em geral, a quantidade de superplastificante a incluir numa composição de betão é determinada experimentalmente por intermédio de ensaios em pastas compostas por cimento, água, uma adição mineral e superplastificante, realizados com o objectivo de determinar a quantidade de superplastificante responsável pela obtenção de uma fluidez máxima da pasta para uma dada relação água/ligante, [10, 62]. Normalmente é suficiente adicionar uns 2% de superplastificante por unidade de peso de cimento. Não obstante, note-se que a maioria dos superplastificantes disponíveis no mercado vêm dissolvidos em água, de modo que a água adicionada à mistura deve ter em conta a água presente no adjuvante. Adicionar demasiada quantidade de superplastificante pode resultar em uma segregação excessiva da argamassa ou betão. Alguns estudos também mostram que demasiado superplastificante pode aumentar o tempo de presa da argamassa ou betão, [72, 90, 155].

A elevada trabalhabilidade obtida na presença de superplastificantes permite a colocação do betão em áreas com pouca acessibilidade ou com elevada densidade de armaduras, assim como facilita a bombagem do betão. Como consequência da redução da razão água/cimento, que pode ser obtida com o uso de superplastificantes, obtêm-se betões com uma maior resistência, impermeabilidade e durabilidade. Os betões recentemente desenvolvidos, como o betão autocompactável e o betão de elevado desempenho, também requerem a presença de superplastificantes.

Os superplastificantes, à semelhança dos plastificantes, são em geral incorporados nos betões com o objectivo de alcançar um dos seguintes efeitos, [51]:

- Aumentar a trabalhabilidade, mantendo constante o consumo de água e por conseguinte a razão água/cimento;
- Reduzir a água de amassadura, mantendo-se a mesma trabalhabilidade, o que, para a mesma dosagem de cimento, reduz a razão água/cimento e eleva as resistências

mecânicas dos betões e aumenta a durabilidade;

- Diminuir o consumo de cimento dos betões, a partir da redução de água e da manutenção da relação água/cimento inicial, sem que haja queda da resistência para uma dada trabalhabilidade.

A principal acção de um superplastificante é a de provocar a dispersão das partículas finas da mistura, nomeadamente de cimento e de adições minerais, de forma a melhorar a fluidez da pasta. Esta dispersão é conseguida por aumento das forças de repulsão resultantes, principalmente, da adsorção de polímeros nas superfícies das partículas de cimento, e de adições minerais eventualmente introduzidas.

Capítulo 3

Procedimento experimental

No presente capítulo apresentam-se as principais propriedades dos materiais seleccionados para a realização do estudo experimental em pastas de cimento destinados a betões auto-compactáveis, bem como os procedimentos empregues nos ensaios adoptados para caracterizar o seu comportamento. Ao longo de todo o trabalho experimental foram adoptados procedimentos rigorosos com o objectivo de eliminar a interferência de factores que não foram considerados no estudo, e que poderiam influenciar os resultados tais como temperatura ambiente, humidade ambiente e temperatura da água usada nas composições, que foi deixada a repousar em recipiente próprio para ter uma temperatura semelhante à ambiental.

O trabalho experimental foi realizado no Laboratório de Betão Estrutural da Faculdade de Engenharia da Universidade do Porto.

As misturas de pastas investigadas neste estudo foram preparadas com cimento, um aditivo mineral, superplastificante e água. As propriedades do cimento e filler calcário usados são apresentados na tabela 3.1. O tamanho médio das partículas de filler calcário é de $4.5\ \mu\text{m}$. Foi usado um superplastificante do tipo solução aquosa de policarboxilatos modificados, Sika Viscocrete 3005®, com as características descritas pela tabela 3.2.

3.1 Preparação da amostra

A mistura de pasta foi preparada em laboratório, num misturador de duas velocidades obedecendo à norma NP EN 196-1 (2006). A sequência de mistura consistiu em adicionar ao cimento e filler calcário, 81% da água e misturar durante 60 segundos, parando de seguida o misturador para limpar as paredes do recipiente, misturando novamente por 60 segundos, e adicionando-se a restante água juntamente com o superplastificante. Em seguida misturou-se por 60 segundos e de seguida limpou-se, novamente a parede do recipiente do material que aderiu a esta. Reinicia-se o misturador por 30 segundos e findo este processo deixou-se repousar por 3 minutos, no fim deste período misturou-se por um período de 30 segundos a uma velocidade mais elevada. Antes de cada um dos ensaios foi agitada a pasta manualmente com uma espátula, com 6 rotações no sentido do ponteiro do relógio, para destruir quaisquer estruturas que se tenham formado durante o repouso, o denominado efeito tixotrópico. A sequência de ensaios foi aproximadamente a seguinte:

- 0h 00 min - mistura da pasta

	CEM I 42.5R	Filler Calcário
C_3S (%)	62	-
C_2S (%)	9.6	-
C_4AF (%)	9.3	-
C_3A (%)	7.6	-
Gesso(%)	4.5	-
$CaCO_3$ (%)	4.6	99.0
Total(%)	97.6	99.0
SiO_2 (%)	18.76	-
Al_2O_3 (%)	5.31	<0.4
Fe_2O_3 (%)	3.23	0.4
CaO (%)	63.19	-
MgO (%)	1.71	-
SO_3 (%)	3.04	-
Perda de Ignição (%)	2.72	43.10
Total (%)	97.96	-
$NaO_{2eq.}$ (%)	0.86	-
CaO livre (%)	1.38	-
Densidade Específica(g/cm ³)	3.16	2.70
Blaine(cm ² /gm)	3270	5150
Resíduo, 90 μm (%)	2.1	7.9
Resíduo, 45 μm (%)	16.9	23.7
Resíduo, 32 μm (%)	29.3	30.8
Tamanho médio(μm)	18.10	8.21

Tabela 3.1: Propriedades químicas e físicas do cimento e filler calcário

- 0h 15 min - início do ensaio do Cone de Marsh
- 0h 20 min - início do ensaio de espalhamento
- 0h 30 min - início do ensaio de centrifugação

Equipamentos utilizados

Foram utilizados os seguintes equipamentos:

- Cone de Marsh e respectivo suporte;
- Centrifugadora da marca Centurion Scientific, modelo K240, série k2;
- mini-cone;
- placa de cerâmica;
- régua;

Viscocrete 3005	
Massa volúmica	$1,05 \pm 0,02 \text{ kg/dm}^3$ (a $+23 \pm 2^\circ\text{C}$)
Teor de sólidos	$26,0 \pm 1,3\%$.
pH ($23 \pm 2^\circ\text{C}$)	$4,0 \pm 1,0$.
Teor de cloretos	$\leq 0,1\%$.

Tabela 3.2: Características segundo informação do fornecedor do superplastificante.

- proveta de 1000 ml;
- gobelé de vidro de 500 ml;
- balança com precisão de $\pm 0,1\%$ da massa a determinar;
- espátula de borracha;
- cronómetro;
- misturadora: as amassaduras das pastas foram efectuadas numa misturadora do tipo prescrito na NP EN 196-1 (2006) que consiste essencialmente numa panela, recipiente em aço inoxidável, e por uma pá misturadora accionada por um motor eléctrico. A pá misturadora roda sobre si mesma, acompanhada de um movimento planetário em torno do eixo do recipiente. Os dois sentidos de rotação são opostos. As velocidades de rotação são as definidas na NP EN 196-1 (2006);
- pano húmido.

3.2 Proporções das misturas

As quantidades dos materiais constituintes da pasta podem ser definidos em função do peso dos sólidos (w_c), da seguinte forma:

$$w_w = (w/c) \times w_c, \quad (3.1)$$

$$w_f = (f/c) \times w_c, \quad (3.2)$$

$$w_{sp} = (sp/c) \times (w_c + w_f). \quad (3.3)$$

Onde w_w é o peso da água, w_f é o peso do filler calcário e w_{sp} é o peso do superplastificante. A quantidade a adicionar por m^3 , descontando a água presente no superplastificante, será:

$$w_{wc} = w_w - w_{sp}(1 - \gamma_{sp}) \quad (3.4)$$

Onde γ_{sp} é o teor de sólidos do superplastificante.

3.3 Caracterização dos materiais utilizados

Ligantes

Utilizou-se um tipo de ligante para a realização das argamassas, o cimento Portland CEM I 42,5R fornecido pela Cimpor - Cimentos de Portugal.

Adições

Recorreu-se a um filer calcário da marca Betocarb® com as suas propriedades descritas na tabela 3.1.

Adjuvantes

O adjuvante escolhido para a realização deste trabalho experimental foi fornecido pela empresa Sika-Portugal e tem a designação comercial de “Sika ViscoCrete 3005®”. É um redutor de água de alta gama com uma base química constituída por carboxilatos modificados e está em conformidade com a especificação do LNEC E374. A massa específica do adjuvante foi fornecida pelo fabricante e a sua descrição é apresentada na tabela 3.2.

Água de amassadura

A água utilizada em todas as amassaduras de pastas foi proveniente da rede pública de abastecimento do Porto. Sendo esta água potável, e de acordo com a especificação do LNEC E372, é considerada apta para o fabrico de betão. Ainda assim, não foi caracterizada por não se considerar relevante a contribuição das características da água utilizada para este trabalho.

3.4 Descrição dos ensaios adoptados

Ensaio do cone de Marsh

O cone de Marsh tem sido usado em diferentes sectores industriais para aferir a fluidez de vários tipos de caldas, nomeadamente na indústria petrolífera e na indústria de construção. Na indústria do cimento é usado como um ensaio de trabalhabilidade para especificação e controle de qualidade de pastas ou caldas de cimento. Pertence à família de ensaios de orifícios como o funil-V ou o ensaio ORIMET. As especificações do ensaio variam de país para país, em Portugal a norma aplicável é a NP EN 445 (2008).

O princípio deste ensaio consiste em preparar uma argamassa ou calda de cimento e medir o tempo que um certo volume demora a escoar através de um funil, dado um determinado diâmetro da abertura do cone. Os cones usados apresentam diversas geometrias e o diâmetro do funil na zona de saída varia entre 5 mm e 12.5 mm. No presente trabalho foi utilizado um funil com uma abertura de 10 ± 1 mm, como definido na norma NP EN 445 (2008). Foram vertidos 1000 ml de pasta no cone, medindo-se o tempo necessário para 500 ml de pasta fluírem pelo funil, doravante esse tempo será designado por T_{flow} e expresso em segundos.

Este tempo de escoamento está ligado à fluidez do material testado, quanto maior o tempo de escoamento menor a fluidez. O tempo de escoamento depende do fluido, mas também é afectado pela geometria do cone. As equações necessárias para resolver o problema de escoamento podem ser derivadas no caso de um fluido seguindo um modelo de Bingham, ver [110, 122], ou pelo modelo de Herschel-Bulkey, ver [109]. Nguyen et. al afirma, em [109], e seguindo a norma francesa NF P18-358 (1985), que se deve usar um cone com um orifício de 10 mm para medir a fluidez de caldas de injeção ou para avaliar a capacidade de fluidificação de adjuvantes.

O procedimento utilizado no ensaio do cone de Marsh é o seguinte:

- Colocar o cone numa posição vertical e segura;
- humidificar o seu interior antes de cada medição;
- colocar uma rede na extremidade onde é vertida a pasta para retirar quaisquer grumos presentes na pasta que possam influenciar a normal fluidez da mesma;
- fechar a extremidade mais fina do cone com um dedo;
- retirar o dedo e iniciar um cronómetro;
- parar o cronómetro quando o recipiente atingir a marca prevista.

Ensaio de centrifugação

O ensaio de centrifugação foi efectuado usando uma centrifugadora da marca Centurion Scientific, modelo K240, série k2. Esta foi usada para compactar os sólidos e para determinar a água livre da pasta, designada por W_{free} , o que permite calcular a densidade de empacotamento da pasta. A água livre é definida como a água que se encontra livre nos espaços vazios entre as partículas. Após 30 minutos do início da preparação da pasta, foram enchidos 4 tubos, com 30 ml em cada recipiente. Foi então efectuada a centrifugação por 15 minutos a 3500 rpm, a uma força de centrifugação relativa de $2500\times g$. No final do processo de centrifugação a água que ficou livre acima da pasta foi retirada com uma pipeta. O peso de cada um dos recipientes foi medido antes da centrifugação, $w_{initial}$, e depois da centrifugação de retirada a água livre, w_{final} . A água livre para cada um dos recipientes foi então calculada usando a seguinte expressão:

$$w_i = \frac{w_{initial} - w_{final}}{30} \quad \text{para } i = 1, 2, 3, 4. \quad (3.5)$$

Sendo que o valor de w_{free} é obtida pela média dos 4 valores anteriores através de:

$$W_{free} = \frac{w_1 + w_2 + w_3 + w_4}{4} \quad (3.6)$$

O procedimento utilizado no ensaio de centrifugação é o seguinte:

- Agitar a pasta com uma espátula para uniformizar o seu conteúdo;
- humidificar o interior de cada um dos recipientes;

- verter 30 ml de pasta em cada um dos recipientes;
- pesar o conteúdo, como atrás foi descrito;
- colocar na centrífugadora e iniciar a mesma;
- no fim do ciclo de centrifugação, retirar os recipientes e destes retirar o excesso de água encontrado na superfície da pasta compactada com uma pipeta;
- pesar novamente cada um dos recipientes e registar o seu peso.

Ensaio de espalhamento

O ensaio de espalhamento foi efectuado para aferir da deformabilidade da pasta e como parâmetro de controle para aceitação dos resultados dos ensaios. Para a realização deste ensaio é necessário dispor de um cone truncado, com diâmetros superior e inferior de 19 e 38 mm, e altura de 57 mm.

Este ensaio consiste em humedecer o interior do cone e a superfície da placa de espalhamento, usando um pano húmido, coloca-se a placa sobre uma superfície firme, plana e nivelada e posicionou-se o cone no centro da mesma, encheu-se o cone de uma só vez, sem qualquer tipo de compactação, nivela-se a pasta no topo do cone e, em seguida, levanta-se de forma cuidadosa e contínua o cone na vertical. No final do movimento da pasta, mediu-se o diâmetro aparentemente máximo da área de espalhamento da pasta, d_{i1} , e o diâmetro perpendicular a este, d_{i2} , onde i representa o número do ensaio. Para o valor do diâmetro de espalhamento da amostra de betão testada tomou-se a média dos dois diâmetros registados, para um melhor controle da amostra, optou-se por efectuar duas amostras, consecutivas, deste ensaio, sendo que d_{1j} designa o diâmetro do primeiro ensaio e d_{2j} o diâmetro do segundo ensaio. O resultado final é calculado usando a seguinte expressão:

$$D_{flow} = \frac{d_{11} + d_{12} + d_{21} + d_{22}}{4} \quad (3.7)$$

3.5 Controlo dos resultados

O tempo de escoamento diminui geralmente com o aumento da dosagem de superplastificante, rácio sp/c . No entanto, para além de uma determinada dosagem o tempo de escoamento não diminui consideravelmente. Esta dosagem particular, tem sido definida como a saturação ou dosagem óptima, ver [4, 74]. A determinação do ponto de saturação é geralmente subjectivo, dependendo da forma da curva do gráfico de dispersão entre o tempo de escoamento e a dosagem de superplastificante e o critério do investigador. O critério aqui usado é o definido por Gomes, em [74], pois a utilização do logaritmo do tempo de escoamento permite a utilização de um critério que depende apenas das características intrínsecas da pasta e não no volume da amostra em si.

Assim para cada rácio w/c foi traçado um gráfico com a dispersão do logaritmo do tempo de escoamento, $\log(T_{flow})$, contra a dosagem de superplastificante, sp/c em %, obtendo-se as Figuras 3.1, 3.2, 3.3, 3.4 e 3.7, onde foi possível aferir em que concentração de

superplastificante terminar a amostra para cada um dos r cios w/c , considerando, tamb m, que a concentra  o m nima de superplastificante a testar seria quando o valor do tempo de escoamento fosse superior a 55 segundos, ou $\log(Tflow) \geq 4.007333$

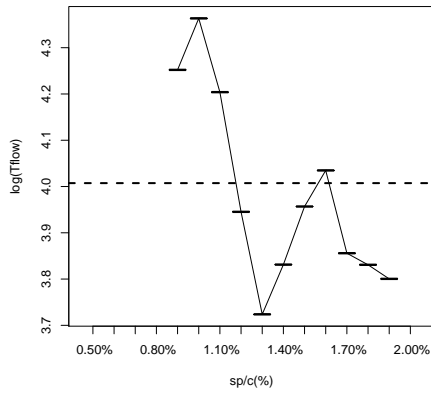
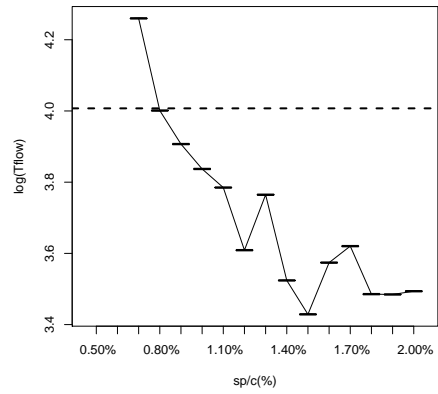
(a) R cio $w/p=0.32$ (b) R cio $w/p=0.33$

Figura 3.1: Procedimentos para a determina  o do ponto de satura  o pelo m todo de AFREM

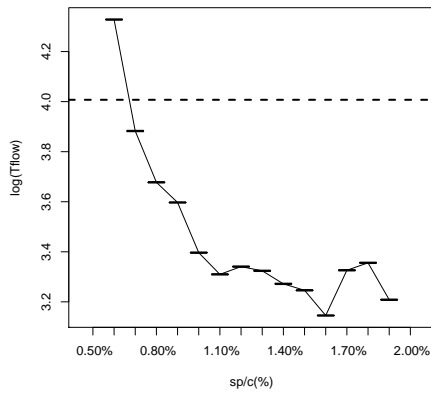
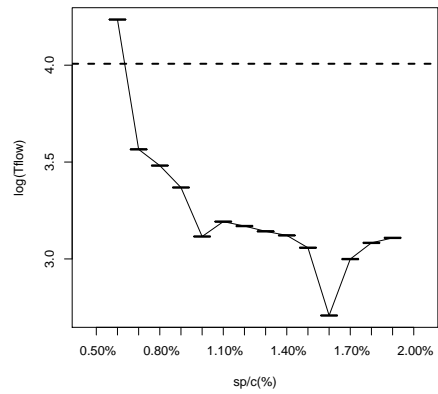
(a) R cio $w/p=0.34$ (b) R cio $w/p=0.35$

Figura 3.2: Procedimentos para a determina  o do ponto de satura  o pelo m todo de AFREM

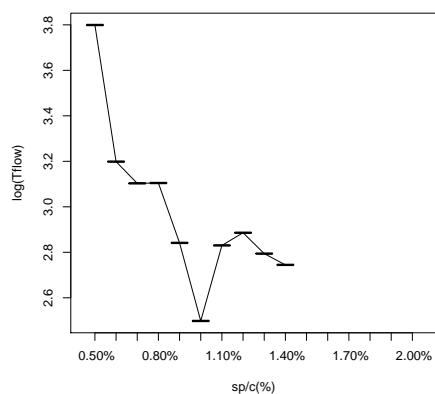
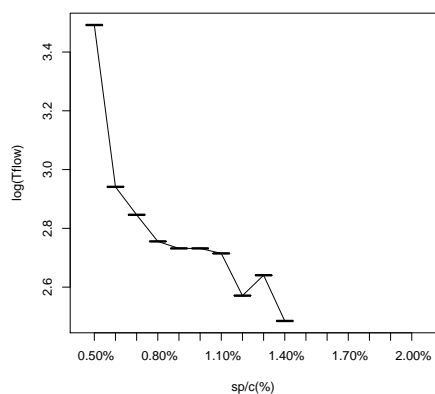
(a) R cio $w/p=0.36$ (b) R cio $w/p=0.37$

Figura 3.3: Procedimentos para a determina  o do ponto de satura  o pelo m todo de AFREM

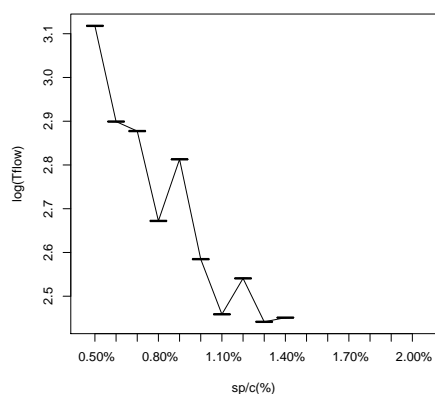
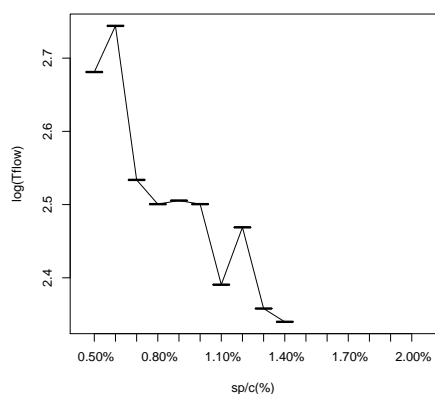
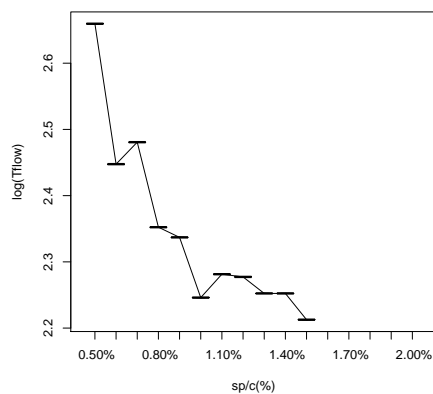
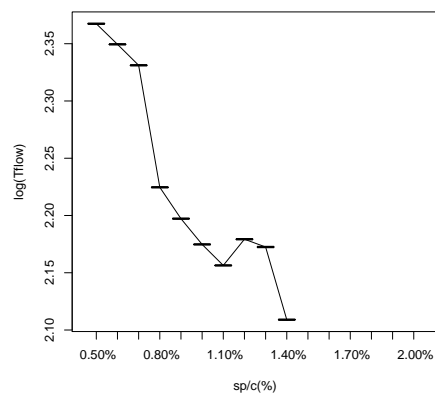
(a) R cio $w/p=0.38$ (b) R cio $w/p=0.39$

Figura 3.4: Procedimentos para a determina  o do ponto de satura  o pelo m todo de AFREM



(a) R cio w/p=0.40



(b) R cio w/p=0.41

Figura 3.5: Procedimentos para a determina  o do ponto de satura  o pelo m todo de AFREM

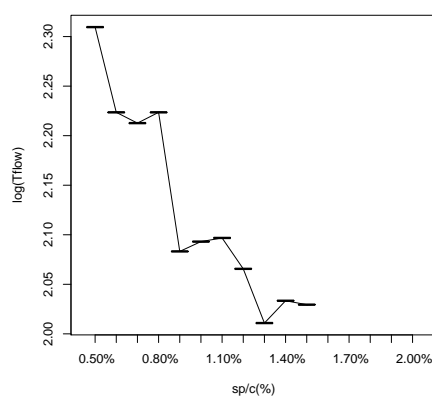


Figura 3.6: R cio w/p=0.42

Figura 3.7: Procedimentos para a determina  o do ponto de satura  o pelo m todo de AFREM

Capítulo 4

Heterocedasticidade dos resíduos da regressão

A regressão linear estuda a relação entre a variável dependente ou variável resposta Y e uma ou várias variáveis independentes ou regressoras X_1, \dots, X_p . O que se procura estabelecer é uma relação funcional entre Y e X_1, \dots, X_p , o que pode ser traduzido por: " Y é explicado em função de X_1, \dots, X_p " ou "os factores explicativos de Y são X_1, \dots, X_p " ver por exemplo [106], num vasto leque de obras introdutórias.

Assim, tem-se a função f de p variáveis:

$$Y = f(X_1, \dots, X_p) \quad (4.1)$$

Parte-se do princípio que a equação (4.1) envolve um conjunto finito de parâmetros β_0, \dots, β_k , tomando esta relação o nome de relação paramétrica.

Um caso particular das relações do tipo descritas por (4.1) são aquelas em que a relação é caracterizada pela linearidade dos parâmetros, isto é, que tomam a seguinte forma:

$$Y = X\beta \quad (4.2)$$

onde β_0, \dots, β_k são os parâmetros, o termo β_0 é o termo independente que é introduzido no modelo e representa o ponto em que a recta ou hiperplano da regressão intercepta o eixo dos yy . Só em casos muito particulares se supõe que a nulidade das variáveis independentes implica a nulidade da variável dependente. Quando assim acontece é necessário impor a restrição $\beta_0 = 0$.

Ocorrem situações em que o modelo (4.1) não é linear relativamente aos parâmetros, mas recorrendo a uma transformação da variável Y , $g(Y)$, é possível, em geral, obter uma relação da forma (4.2).

Quando se considera 4.2 assume-se que Y é apenas explicado por X . Contudo em situações reais essa hipótese ideal é muitas vezes contrariada, pois devido à natureza dos dados existem variáveis que não são observáveis ou também existem erros de observação e lidamos muitas vezes com quantidades aleatórias. O que implica que seja rígida e portanto seja necessário introduzir uma variável adicional, que se designara por ϵ .

Se ϵ for inserido no modelo aditivamente, o modelo teórico definido passa a ser:

$$Y = X\beta + \epsilon \quad (4.3)$$

Como ϵ não é observável, é designado por resíduo, erro ou perturbação. Desta forma o modelo (4.3) tem duas componentes, a componente sistemática ou sinal, $X^\top\beta$ e a componente residual ou ruído ϵ .

Exige-se fazer também uma distinção entre tipos de dados. Dados seccionais são observações de várias variáveis numa mesma data ou período de tempo. Enquanto que dados longitudinais são fruto de observações de uma mesma variável em várias datas.

Considere-se um conjunto de dados com n observações. Então define-se

$$Y = (Y_1, Y_2, \dots, Y_n)^\top$$

como o vector de observações da variável dependente Y . Defina-se a seguinte matriz

$$X = \begin{pmatrix} 1 & x_{11} & x_{21} & \cdots & x_{1k} \\ 1 & x_{12} & x_{22} & \cdots & x_{1k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{nk} \end{pmatrix}$$

em que a primeira coluna representa o termo independente, em que cada uma das restantes colunas representa uma variável independente. Defina-se também os seguintes vectores de parâmetros

$$\beta = (\beta_0, \beta_1, \dots, \beta_k)^\top$$

e o vector de resíduos

$$\epsilon = (\epsilon_0, \epsilon_1, \dots, \epsilon_k)^\top$$

Como é amplamente conhecido um dos pressupostos do modelo de regressão linear (4.3) é o de que os erros devem ter variância constante. Esta condição é designada por homocedasticidade. A variância ser constante equivale a supor que não existem observações incluídas na variável residual cuja influência seja mais intensa na variável dependente. Uma das técnicas usadas para verificar a suposição de que os resíduos são homoscedásticos, é a análise do gráfico dos resíduos versus valores ajustados. Este gráfico deve apresentar pontos dispostos aleatoriamente sem nenhum padrão definido, como se pode ver, por exemplo na Figura 4.1. Se os pontos estão aleatoriamente distribuídos em torno da recta $y = 0$, sem nenhum comportamento ou tendência, temos indícios de que a variância dos resíduos é constante. Já a presença, por exemplo, de uma forma em funil é um indicativo da presença de heterocedasticidade. A hipótese de igualdade de variância, ou homocedasticidade, é fundamental na análise de modelos de regressão clássicos [39]. Apesar de, teoricamente, esta suposição ser importante, não é satisfeita em algumas situações [145, 152]. Variância não constante dos resíduos do modelo, ou heterocedasticidade, é encontrada em muitos campos de investigação como, apenas para citar alguns, biologia [39], desenho de experiências [22, 23] e muito frequentemente em econometria [7, 17, 39, 69, 97]. É o caso também do exemplo de aplicação que nos propomos abordar no Capítulo 5 em que tratámos o modelo

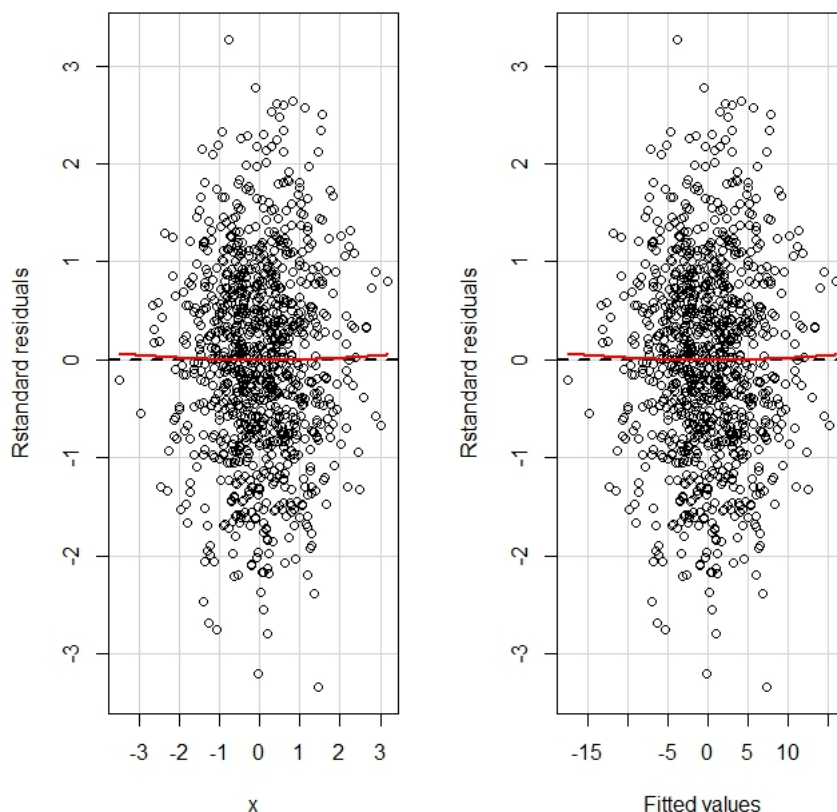


Figura 4.1: Gráfico de dispersão dos resíduos estudentizados contra a variável independente e contra os valores ajustados do modelo

de propriedades de pastas de betão.

Existem vários tipos de heterocedasticidade. Pode acontecer por exemplo que um conjunto de dados sejam heterocedástico, mas se dividido em subgrupos de dados, estes já manifestam uma relação homocedástica. Outro tipo de heterocedasticidade é aquela em que a variância depende da média da variável resposta ou de uma ou mais variáveis explicativas.

A heterocedasticidade e o seu impacto no modelo de regressão linear gaussiano tem sido profundamente estudado, especialmente em econometria. Ficou suficientemente demonstrado que, usando regressão linear para modelar uma relação existente entre variáveis e fazer previsões com dados heterocedásticos, pode resultar em falhas graves, pois a região com baixa variância tem menor influência na determinação dos parâmetros e na determinação de intervalos de confiança, do que regiões com maior variância, [12, 69, 97, 115].

A modelação da variância pode, na regressão linear com dados heterocedásticos, atingir erros de Tipo I e melhorar as probabilidades dos intervalos de confiança, para previsões baseadas em modelos de regressão gaussianos, permitindo alcançar o nível de potência nominal [39, 145].

Considere-se o modelo de regressão heterocedástico

$$\begin{aligned} Y &= X\beta + \epsilon \\ \text{var}(\epsilon_i|x_i) &= \sigma_i^2, \quad i = 1, \dots, n. \end{aligned} \quad (4.4)$$

Continuamos a assumir que as perturbações não são correlacionadas duas a duas, ou seja, a matriz de covariância ϵ é diagonal

$$\mathbb{E}[\epsilon\epsilon^\top] = \sigma^2\Omega = \sigma^2 \begin{pmatrix} \omega_1 & 0 & 0 & \dots & 0 \\ 0 & \omega_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \omega_n \end{pmatrix}$$

Existem dois tipos de abordagem para estabilizar a heterocedasticidade [39]:

- por via de transformações apropriadas;
- usando o método dos mínimos quadrados ponderados.

Nas secções seguintes vai-se descrever os efeitos que a heterocedasticidade têm na determinação dos parâmetros de regressão e na análise estatística do modelo obtido e discutir-se métodos para aferir da presença da heterocedasticidade e para a sua modelação.

4.1 Consequências de ignorar a heterocedasticidade

Os resultados anteriores implicam que o vector $\hat{\beta}$ obtido pelo método dos mínimos quadrados, na presença de heterocedasticidade, é ineficiente relativamente ao estimador obtido pelo método dos mínimos quadrados ponderados. Outras dificuldades que advêm de ignorar a heterocedasticidade patente nos dados são a seguir brevemente aforadas e comentadas.

Ineficiência do Estimador de Mínimos Quadrados

Se a matriz de covariância do estimador $\hat{\beta}$ do método dos mínimos quadrados, a saber

$$\Sigma_{\hat{\beta}} = \sigma^2(X^\top X)^{-1}, \quad (4.5)$$

e a matriz de covariância do estimador do método dos mínimos quadrados ponderados,

$$\Sigma_{\hat{\beta}} = \sigma^2(X^\top X)^{-1}(X^\top \Omega X)(X^\top X)^{-1} \quad (4.6)$$

não coincidem então o estimador do desvio padrão dos parâmetros é desadequado.

Efeitos nos testes de hipóteses

As estimativas das variâncias e covariâncias do estimador de β pelo método dos mínimos quadrados são enviesados e inconsistentes quando a heterocedasticidade está presente (ver [94, 153]), logo os testes de hipóteses são inválidos. Basta notar que, para o teste F temos

que como a variância estimada, s^2 , não é constante, o teste é habitual perde sentido e precisaria de ser reformulado.

Efeitos na previsão

Previsões baseadas no método dos mínimos quadrados são não enviesadas, mas as previsões são ineficientes, porque as estimativas são ineficientes. Destacamos as seguintes particularidades dos estimadores do método dos mínimos quadrados na presença de heterocedasticidade:

- como não são verificadas as condições de Gauss-Markov não temos garantia de que o estimador pelo método dos mínimos quadrados seja, em geral, o melhor estimador linear não enviesado e que seja eficiente. As previsões obtidas por este estimador são também, em geral, ineficazes;
- os desvios padrão estimados e covariâncias dos coeficientes de regressão são tendenciosos e inconsistentes, e, portanto, os intervalos de confiança, seja para os parâmetros seja para as previsões, são inválidos.

Matriz de covariância do estimador dos parâmetros

Seja

$$s^2 = \frac{\epsilon^\top \epsilon}{n - k} = \frac{\hat{\epsilon}^\top M \hat{\epsilon}}{n - k} \quad (4.7)$$

onde $M = I_n - X(X^\top X)^{-1}X^\top$. Logo, substituindo a expressão de M em (4.7) temos

$$s^2 = \frac{\hat{\epsilon}^\top X(X^\top X)^{-1}X^\top \hat{\epsilon}}{n - k} \quad (4.8)$$

o que implica que a diferença, em termos de segundos momentos, entre o estimador usual e o estimador apropriado é [69]:

$$\text{var}(\hat{\beta}|X) - \text{var}(\hat{\hat{\beta}}|X) = \sigma^2(X^\top X)^{-1} - \sigma^2(X^\top X)^{-1}(X^\top \Omega X)(X^\top X)^{-1}. \quad (4.9)$$

Numa amostra suficientemente grande, tal que $s^2 \approx \sigma^2$, a diferença entre variâncias é aproximadamente

$$\frac{\sigma^2}{n} \left(\frac{X^\top X}{n} \right)^{-1} \left(\frac{X^\top X}{n} - \frac{X^\top \Omega X}{n} \right) \left(\frac{X^\top X}{n} \right)^{-1} \quad (4.10)$$

e daqui temos que essa diferença é

$$\Delta = \frac{X^\top X}{n} - \frac{X^\top \Omega X}{n} = \sum_{i=1}^n \left(\frac{1}{n} \right) x_i x_i^\top - \sum_{i=1}^n \left(\frac{w_i}{n} \right) x_i x_i^\top, \quad (4.11)$$

onde se considera que os pesos são 1 para o primeiro estimador e w_i para o segundo estimador. A diferença entre os pesos w_i/n e $1/n$ depende dos w_i usados. Se estes são relacionados com x_i a diferença pode ser considerável, caso contrário, se os pesos não são correlacionados com $x_i x_i^\top$, os pesos tenderão a ser iguais. Portanto, a eficiência do estimador $\hat{\beta}$ prende-se com o facto de a heterocedasticidade ser ou não relacionada com x_i ou $x_i \times x_j$. Logo a conclusão

a tirar é que, em geral, se a heterocedasticidade não for correlacionada com as variáveis do modelo, pelo menos em grandes amostras, o estimador dos mínimos quadrados, apesar de não ser eficiente, não é incorrecto. Quando se está numa situação em que o estimador do método mínimos quadrados for ineficiente deve-se usar o método dos mínimos quadrados ponderados que é um caso particular do método dos mínimos quadrados generalizado que a seguir se introduz.

4.2 Método dos mínimos quadrados generalizados

O modelo linear gaussiano têm como hipótese a variância constante, ou seja,

$$\mathbb{E}(\epsilon_i|X) = \sigma^2 > 0 \quad \text{para } i = 1, \dots, n.$$

No entanto, em muitas situações práticas, isso não é possível. Então torna-se necessário relaxar essa hipótese. Como tal, vamos supor que

$$\text{var}(Y) = \sigma^2 \Omega \quad (4.12)$$

onde Ω é uma matriz diagonal, $n \times n$, conhecida. A interpretação desta condição é que, se Ω é diagonal, então as observações de Y não são correlacionadas, mas têm variâncias diferentes. Se algum elemento de Ω pertencente à diagonal fosse não nulo então as observações seriam correlacionadas.

Considere-se o seguinte modelo

$$\begin{aligned} Y &= X\beta + \epsilon \\ \mathbb{E}(\epsilon) &= 0 \\ \text{var}(\epsilon) &= \sigma^2 \Omega \end{aligned} \quad (4.13)$$

então o estimador $\hat{\beta}$ determinado pelo do método dos mínimos quadrados, já não é eficiente. Como $\sigma^2 \Omega$ é a matriz de covariância dos erros, Ω é não singular e definida positiva, logo existe uma matriz U , $n \times n$, não singular e simétrica, tal que $\Omega = U^T U = U U^T$.

Defina-se então

$$\begin{aligned} Z &= U^{-1}Y \\ V &= U^{-1}X \\ u &= U^{-1}\epsilon \end{aligned} \quad (4.14)$$

O modelo (4.13) vem

$$U^{-1}Y = U^{-1}X\beta + U^{-1}\epsilon \quad (4.15)$$

ou seja

$$Z = V\beta + u. \quad (4.16)$$

Da hipótese exogeneidade do método dos mínimos quadrados vem que os erros deste último modelo são centrados $\mathbb{E}(u) = 0$. Mais, a matriz de covariância de u é dada por

$$\begin{aligned}
 \text{var}(u) &= \mathbb{E}((u - \mathbb{E}(u))(u - \mathbb{E}(u))^T) \\
 &= \mathbb{E}(uu^T) \\
 &= \mathbb{E}(U^{-1}\epsilon\epsilon^T U^{-1}) \\
 &= U^{-1}\mathbb{E}(\epsilon\epsilon^T)U^{-1} \\
 &= \sigma^2 U^{-1}\Omega U^{-1} \\
 &= \sigma^2 I
 \end{aligned}$$

Logo os elementos de u tem média nula e variância constante. Como estamos nas condições do modelo de regressão linear mais simples, podemos aplicar o método dos mínimos quadrados isto é, minimizamos

$$\begin{aligned}
 SQR &= \sum_{i=1}^n u_i^T u_i = u^T u \\
 &= (Y - X\beta)^T \Omega^{-1} (Y - X\beta) \\
 &= Y^T \Omega^{-1} Y - 2\beta^T X^T \Omega^{-1} Y + \beta^T X^T \Omega^{-1} X \beta.
 \end{aligned} \tag{4.17}$$

Derivando em ordem a β , obtemos

$$\frac{\partial SQR}{\partial \beta} = -2X^T \Omega^{-1} Y + X^T \Omega^{-1} X \beta \tag{4.18}$$

que tem como solução

$$\hat{\beta} = (X^T \Omega^{-1} X)^{-1} X^T \Omega^{-1} Y. \tag{4.19}$$

O estimador apresentado em (4.19) deve ser usado na prática sempre que os pressupostos do modelo de regressão linear pelo método dos mínimos quadrados usual sejam violados. A dificuldade da sua utilização está no facto de Ω ser suposto conhecido. Em termos estatísticos a ideia será substituir Ω por uma sua estimativa empírica com os perigos que daí advêm para as propriedades do estimador. Para tal considere-se o método dos mínimos quadrados ponderados a seguir descrito.

4.3 Método dos mínimos quadrados ponderados

Um dos pressupostos subjacentes à maioria dos métodos de modelação matemática de regressão por mínimos quadrados, incluindo lineares e não lineares, é que cada ponto correspondente a uma observação fornece informações igualmente precisas sobre a parte determinística da variação total do processo, ou seja, o desvio padrão do termo de erro é constante ao longo de toda a distribuição das variáveis explicatórias. Esta hipótese, no entanto, claramente não é sustentável em todas as aplicações. Em situações como a que acabamos de descrever, quando não parece ser razoável supor que cada observação deve ser tratada de igual forma, o método dos mínimos quadrados ponderados pode muitas vezes

ser utilizado para maximizar a eficiência na estimação de parâmetros. Isso é feito para tentar dar a cada indivíduo observado a sua quantidade adequada de influência sobre as estimativas dos parâmetros. Um procedimento que trata todos os dados igualmente daria a pontos medidos com menos precisão mais influência do que deveriam ter e daria a pontos de elevada precisão muito pouca influência, ver [39].

Quando os erros, ϵ , não são correlacionados, mas têm variância não constante, tal que a matriz de covariância de ϵ é

$$\sigma^2 \Omega = \sigma^2 \begin{pmatrix} \frac{1}{w_1} & 0 & \dots & 0 \\ 0 & \frac{1}{w_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{w_n} \end{pmatrix} \quad (4.20)$$

Considerando

$$W = \Omega^{-1},$$

como Ω é uma matriz diagonal, W também é uma matriz diagonal cujos elementos são os pesos w_1, w_2, \dots, w_n . Substituindo Ω^{-1} por W em (4.19), temos o estimador de mínimos quadrados ponderados de β :

$$\hat{\beta} = (X^\top W X)^{-1} X^\top W Y. \quad (4.21)$$

Seja, então, $Z = z_1, \dots, z_n$ e $h(Z) = \text{diag}(W)$

$$\text{var}(\epsilon|Z) = \sigma^2 w(Z) \quad (4.22)$$

onde $w(Z) > 0$, com vista a preservar a positividade do segundo momento, é uma função da variável que representa a heterocedasticidade, assumindo que $w(Z)$ é conhecida, temos que para cada uma das observações i da população, podemos expressar a variância dos parâmetros na equação de regressão como

$$\sigma_i^2 = \text{var}(\epsilon|z_i) = \sigma^2 w(z_i) = \sigma^2 w_i \quad (4.23)$$

onde w_i varia com cada uma das observações z_i . Dada essa estrutura, pode-se então transformar a equação de regressão dos mínimos quadrados, de modo que o erro se torne homocedástico e satisfaça os pressupostos do teorema de Gauss-Markov.

Considere-se a transformação onde se divide o termo do erro por $\sqrt{w_i}$. Então o novo erro fica ϵ_i/w_i . Com isto temos que a variância do modelo fica

$$\text{var} \left(\frac{\epsilon_i}{\sqrt{w_i}} | z_i \right) = \mathbb{E} \left(\frac{\epsilon_i^2}{h_i} | z_i \right) = \sigma^2 \frac{w_i}{w_i} = \sigma^2. \quad (4.24)$$

Então o novo termo é homocedástico, e como consequência disto podemos dividir a equação de regressão por $\sqrt{h_i}$, tal que

$$\frac{Y}{\sqrt{w(Z)}} = \frac{X^\top \beta}{\sqrt{w(Z)}} + \frac{\epsilon}{\sqrt{w(Z)}} \quad (4.25)$$

Embora esta nova equação de regressão parecer peculiar, e parecer complicar a interpretação, o investigador tem que ter em mente que esta transformação é realizada de modo a alcançar a eficiência dos parâmetros, ou seja, obter a variância correta. Quando se interpretam os resultados, o investigador deve estar preocupado com a forma de regressão original, não transformada, e não a equação de regressão pelo método dos mínimos quadrados ponderados. Por fim, note-se que a equação (4.25) ainda é linear nos parâmetros, e satisfaz todas os pressupostos do teorema de Gauss-Markov para o método dos mínimos quadrados, desde que além da heterocedasticidade, todos os outros pressupostos não sejam violados.

Uma observação pertinente sobre este método é que atribui maior peso a observações com menor variância, ou seja, as observações com menor variância tem um contributo maior para a determinação das estimativas dos parâmetros,. No entanto o estimador obtido por esta via apresenta boas propriedades, sendo uma delas fulcral para os nossos propósitos, que vem do facto de, ver [144], Víšek (2011) ter demonstrado a sua consistência e robustez mesmo na presença de heterocedasticidade.

Em termos de implementação do método, existem versões iterativas, com grande interesse prático.

Versão iterativa dos mínimos quadrados ponderados

Considere-se a norma L_p , onde $1 < p < 2$, representada por:

$$\|Y - X^\top \hat{\beta}\|^p = (Y - X^\top \hat{\beta})^\top W (Y - X^\top \hat{\beta}) \quad (4.26)$$

onde

$$W = \text{diag}(|y_i - x_i^\top \hat{\beta}|^{2-p}),$$

sendo x_i e y_i a linha i de X e de Y respectivamente. Descrevemos o seguinte algoritmo iterativo para o método dos mínimos quadrados ponderados, em que a sequência de pesos $W^{(k)}$ e coeficientes $\hat{\beta}^{(k)}$ são encontrados passo-a-passo, ver [86].

1. Definir um limite para o resíduo inicial, ε_1 , que pode ser dependente dos dados; definir um valor elevado para o peso dos resíduos W_{big} ; definir critérios de paragem, ε_2 e k_{max} .

2. Calcular a matriz diagonal $W^{(k)}$ tal que

$$W_i^{(k)} = |y_i - x_i^\top \hat{\beta}^{(k)}|^{2-p},$$

excepto se $|y_i - x_i^\top \hat{\beta}^{(k)}|^{2-p} < \varepsilon_1$. Tomar $w_i = w_{big}$.

3. Calcular $\hat{\beta}^{(k)}$ resolvendo o problema do método dos mínimos quadrados ponderado,

isto é minimizando

$$(Y - X^\top \hat{\beta})^\top W (Y - X^\top \hat{\beta}).$$

4. Se $\|\hat{\beta}^{(k+1)} - \hat{\beta}^{(k)}\| \leq \varepsilon_2$, tomar $\hat{\beta} = \hat{\beta}^{(k+1)}$ e terminar.
5. Se $k < k_{max}$, voltar ao passo (2); caso contrário o algoritmo não convergiu no número de passos pretendido.

Este método é de fácil implementação e fiável, excepto quando o valor absoluto dos resíduos for igual ou muito próximo de zero, para mais pormenores ver [86].

Coeficiente de determinação em regressão pelo método dos mínimos quadrados ponderados

Considere-se o modelo

$$Y = X^\top \beta + \epsilon \quad (4.27)$$

onde β é um vector, $k \times 1$, de parâmetros desconhecidos, ϵ é um vector, $n \times 1$, de variáveis aleatórias não observadas, com $\epsilon \sim \mathcal{N}(0, \sigma^2 W)$, e W é uma matrix diagonal, $n \times n$, com elementos $w_{ii} > 0$ na sua diagonal. Como usual W não é conhecida e é necessário estimar através dos resíduos de uma regressão pelo método dos mínimos quadrados.

Seja, então a equação transformada

$$W^{-1/2}Y = W^{-1/2}X^\top \beta + W^{-1/2}\epsilon \quad (4.28)$$

fazendo $Y^* = W^{-1/2}Y$, $X^* = W^{-1/2}X^\top$ e $\epsilon^* = W^{-1/2}\epsilon$, de onde obtemos

$$Y^* = X^* \beta + \epsilon^*. \quad (4.29)$$

Willett et. al (ver [154]) afirma que o coeficiente de determinação de uma regressão pelo método dos mínimos quadrados ponderados não tem a mesma utilidade que o coeficiente de determinação usual, pois

$$R_{WLS}^2 = 1 - \frac{SQR}{SQT} = 1 - \left(\frac{(Y^* - X^* \hat{\beta}^*)^\top (Y^* - X^* \hat{\beta}^*)}{Y^{*\top} Y^* - n \bar{Y}^{*2}} \right) \quad (4.30)$$

onde $\hat{\beta}^*$ são os coeficientes de regressão pelo método dos mínimos quadrados ponderados. Logo R_{WLS}^2 é o coeficiente de determinação dos dados transformados, e é a medida de proporção da variação de Y ponderado explicado por X ponderado. Assim o que o coeficiente de determinação que surge na maioria dos pacotes estatísticos computacionais quando é feita uma regressão pelo método dos mínimos quadrados ponderados reflecte a eficiência da ponderação em resolver o problema de heterocedasticidade: mas de um ponto de vista aplicado é mais realista e eficiente reportar o coeficiente na métrica original. Então Willett et.al. propõe um novo coeficiente de determinação

$$pseudo \quad R_{WLS}^2 = 1 - \left(\frac{(Y - X^\top \hat{\beta}^*)^\top (Y - X^\top \hat{\beta}^*)}{Y^\top Y - n \bar{Y}^2} \right) \quad (4.31)$$

uma estatística que obviamente será menor que a obtida na regressão original sem ponderação. A utilidade de R^2 para a inferência da qualidade de ajustamento de um modelo de regressão tem sido debatido ao longo dos tempos, ver [28, 29, 96, 100], como complemento a este valor foi sugerida a estatística \bar{R}^2 , que se revela mais fiável. Draper, em [56], propõem um outro critério, apelidado de critério Box-Wetz, como alternativa a R^2

$$u = \frac{\max \hat{Y} - \min \hat{Y}}{\sqrt{(k-1)s^2/n}} \quad (4.32)$$

onde \hat{Y} são os valores ajustados da regressão, s^2 estima a variância do modelo, σ^2 , e k é o número de parâmetros do modelo. Draper considera que um bom ajustamento do modelo é obtido para valores de u superiores a 10.

Perante estes resultados conclui-se que basear qualquer análise estatística de um modelo de regressão linear, tanto clássico como ponderado, usando apenas o coeficiente de determinação para avaliar da qualidade do ajustamento, impedirá a revelação de características importantes dos dados e inadequação do modelo, [96], sendo que Kvalseth recomenda uma análise detalhada dos resíduos para revelar estruturas e informação que poderão melhorar a qualidade da análise do modelo.

4.4 Resíduos

Todos nós sabemos que os resíduos são a base dos métodos de diagnóstico em modelos de regressão linear. Suponha-se que ajusta um modelo aos dados, assumindo uma variância do erro constante σ^2 . Os resíduos comuns são dados pela diferença entre as respostas e os valores ajustados

$$\epsilon_i = y_i - \hat{y}_i, \quad \text{para } i = 1, \dots, n \quad (4.33)$$

Os resíduos comuns não são correlacionados com os valores ajustados ou quaisquer combinações lineares das variáveis independentes. Logo, padrões que se identifiquem nos gráficos dos resíduos contra qualquer uma das variáveis revelam violações das hipóteses do modelo.

Se o modelo estiver devidamente ajustado, então os resíduos comuns são variáveis aleatórias com média nula e variância dada por:

$$\text{var}(\epsilon_i) = \sigma^2(1 - h_i) \quad (4.34)$$

onde h_i é chamado de valor de alavanca. Em modelos lineares este valor está compreendido entre 0 e 1 (ver [147]) dependendo da localização dos valores previstos para uma determinada observação em relação a outras observações. Valores levados de h_i correspondem a observações com valores invulgares de x_i que puxam a superfície de regressão, enquanto valores baixos de h_i correspondem a observações próximas do modelo considerado.

Resíduos de observações com valores elevados de h_i tem menor variância. Para corrigir este efeito pode-se dividir o resíduo por uma estimativa do seu desvio padrão e assim surgem os resíduos studentizados:

$$\epsilon_{Ti} = \frac{\epsilon_i}{\hat{\sigma}(i)} \sqrt{1 - h_i} \quad (4.35)$$

onde $\hat{\sigma}_{(i)}^2$ é uma estimativa de σ^2 calculado da regressão eliminando a observação i . Se os erros originais forem normalmente distribuídos, então ϵ_{Ti} segue uma distribuição t de Student com $n - k - 2$ graus de liberdade e os resíduos estudentizados podem ser usados para testar outliers.

Se o modelo é ajustado usando o método dos mínimos quadrados ponderados, com pesos w_i , conhecidos e positivos, então os resíduos comuns são substituídos pelos resíduos de Pearson

$$\epsilon_{Pi} = \sqrt{w_i} \epsilon_i \quad (4.36)$$

Os resíduos estudentizados não são afectados pelos pesos, pois os pesos são cancelados pelo numerador e denominador na respectiva formula. Para além dos habituais testes para a análise da gaussianidade dos resíduos, da sua independência e colinearidade é recomendado a análise gráfica dos mesmos.

Gráficos de diagnóstico

Cook et. al [44, 45] sugerem que em complemento a testes formais utilizados para diagnosticar a violação da hipótese nula de homocedasticidade do modelo de regressão gaussiano, se use análise de gráficos dos resíduos contra as variáveis independentes ou os valores ajustados.

Considere-se então o modelo

$$\begin{aligned} Y &= X^\top \beta + \epsilon \\ \epsilon &\sim \mathcal{N}(0, \sigma^2 I) \end{aligned} \quad (4.37)$$

onde β é um vector, $k \times 1$, de parâmetros desconhecidos, ϵ é um vector, $n \times 1$, de variáveis aleatórias não observadas, e σ^2 , constante desconhecida, é a variância do modelo (4.37). Colocando a hipótese alternativa de que a variância não é constante, então a matriz de covariância dos erros é

$$\sigma_0^2 W, \quad (4.38)$$

onde W é como na Secção 4.3 e σ_0^2 é o verdadeiro parâmetro de escala ou variância do modelo. A matriz W é uma matriz diagonal, $n \times n$, com $w_i = w(z_i, \delta)$ onde $z_i^\top = (z_i 1, \dots, z_i q)$ é a i -ésima linha de uma matriz conhecida Z , Z é a matrix de parte ou totalidade das variáveis independentes X , sendo δ um vector de dimensão $q \times 1$.

Se ajustarmos o modelo (4.37) ao conjunto de dados, mas a matriz de covariância dos verdadeiros erros é (4.38) então

$$\begin{aligned} \mathbb{E}(\hat{\epsilon}_i | z_i) &= 0 \\ \text{var}(\hat{\epsilon}_i | z_i) &= \mathbb{E}(\hat{\epsilon}_i^2 | z_i) = ((1 - h_{ii})^2 w_i + \sum_{j \neq i} h_{ij}^2 w_j) \sigma_0^2 \end{aligned} \quad (4.39)$$

Como $\mathbb{E}(\hat{\epsilon}_i^2 | z_i)$ depende de $(1 - h_{ii})$ mesmo quando $W = I$, Cook et. al [44] sugerem substituir $\hat{\epsilon}_i$ por $b_i = \hat{\epsilon}_i / \sqrt{(1 - h_{ii})}$, onde $\hat{\epsilon}_i$ é o resíduo i do modelo (4.37).

Apresentamos de seguida uma breve discussão dos diferentes tipos de gráficos de resíduos a que podemos recorrer.

Gráficos lineares dos resíduos

Se se suspeitar de que a variância é não constante e é uma função de variável peso z_i , um método de diagnóstico consiste em construir o gráfico de dispersão da variável b_i contra a variável z_i , pois a média de $b_i|z_i$ é zero (4.39). Logo a forma deste gráfico é determinado pela variância:

$$\text{var}(b_i|z_i) = \mathbb{E}(b_i^2|z_i) = \left((1 - h_{ii})w_i + \sum_{j \neq i} \frac{h_{ij}^2 w_j}{1 - h_{ii}} \right) \sigma_0^2 \quad (4.40)$$

Para melhor se observar a função peso, w , é sugerido que se remova o factor $(1 - h_{ii})$ do primeiro termo da equação (4.40). Isso é conseguido fazendo o gráfico da variável $\hat{\epsilon}_{(-i)} = b_i / \sqrt{(1 - h_{ii})}$ contra a variável z_i , onde $\hat{\epsilon}_{(-i)}$ são os resíduos estudentizados do modelo de regressão obtido pelo método dos mínimos quadrados. Se a variância for, por outro lado, uma função da resposta esperada, então as expressões (4.39) indicam que a forma do gráfico linear dos resíduos é dado por

$$\mathbb{E}(\hat{\epsilon}_i|\hat{y}_i) \pm s.e.(\hat{\epsilon}_i|\hat{y}_i).$$

Como a média condicional, $\mathbb{E}(\hat{\epsilon}_i|\hat{y}_i)$, não é completamente conhecida, pois $x_i^\top \beta$ é desconhecido, substituindo $x_i^\top \beta$ por $x_i^\top \hat{\beta}$, então a forma do gráfico pode ser aproximado considerando $0 \pm s.e.(\hat{\epsilon}_i|\hat{y}_i)$ contra \hat{y}_i . Como exemplo ilustrativo considere-se a Figura 4.2.

Gráficos quadráticos dos resíduos

Como os gráficos dos resíduos lineares conduzem a alguma dificuldade para identificar a função w directamente, pois, repare-se na Figura 4.2, esses resíduos são simétricos em relação ao eixo horizontal, para eliminar essa dificuldade Tsai et. al [138], recomendam que se apresente o gráfico dos resíduos quadrados $\hat{\epsilon}_i^2$ contra z_i . O padrão deste gráfico é determinado por

$$\mathbb{E}(\hat{\epsilon}_i^2|z_i) \pm s.e.(\hat{\epsilon}_i^2|z_i)$$

onde $s.e.(\hat{\epsilon}_i^2|\hat{z}_i) = \sqrt{2\text{var}(\hat{\epsilon}_i^2|z_i)}$. Obviamente o limite inferior deste gráfico é 0 pois os resíduos quadrados são não negativos. Se o segundo termo na equação (4.40) for ignorado, então $s.e.(\hat{\epsilon}_i^2|\hat{z}_i) = \sqrt{2}w_i$. Logo o padrão do gráfico dos resíduos estudentizados pode identificar directamente a função peso w para situações de variância não constante. Note-se que o mesmo se aplica para os valores ajustados da forma que foi referida quando tratámos da aproximação em gráficos lineares de resíduos. A Figura 4.3 ilustra esta abordagem.

Gráfico parcial de resíduos

Um gráfico parcial de resíduos é usado para visualizar os efeitos de uma variável independente na resposta, ao remover os efeitos das outras variáveis independentes. O resíduo parcial para a variável j , para $j = 1, \dots, k$, é

$$Y_i - \left(\hat{\beta}_0 + \sum_{j' \neq j} X_{ij'} \hat{\beta}_{j'} \right) = \hat{Y}_i + \hat{\epsilon}_i - \left(\hat{\beta}_0 + \sum_{j' \neq j} X_{ij'} \hat{\beta}_{j'} \right) = X_{ij} \hat{\beta}_j + \hat{\epsilon}_i. \quad (4.41)$$

De notar que o lado esquerdo de (4.41) indica que o resíduo parcial é a resposta com os efeitos de todas variáveis independentes, excepto a variável j . O lado direito de (4.41) indica que o resíduo parcial é também igual ao resíduo com o efeito da variável j adicionado de volta. O gráfico parcial de resíduos é simplesmente o gráfico de dispersão da resposta contra estes resíduos parciais. Vejamos o seguinte exemplo ilustrativo.

Considere-se o modelo

$$Y = 5X + 2X^2 + \epsilon, \quad (4.42)$$

onde $X \sim \mathcal{N}(0, 1)$ e $\epsilon \sim \mathcal{N}(0, 1)$, e ajuste-se o modelo linear

$$Y = X, \quad (4.43)$$

usando o método dos mínimos quadrados. Como se pode verificar pela Figura 4.4 o nosso modelo linear ajustado (linha vermelha) não constitui um bom ajuste aos dados. Pelo contrário a linha verde, obtida usando uma regressão por Loess, (ver [85]) para mais detalhes sobre este método e outros métodos não paramétricos), traduz um ajustamento melhor aos dados, ou seja, o modelo (4.43) não reflecte o comportamento das observações e portanto é necessário adicionar um termo ao modelo (4.43), sendo que neste caso o termo é X^2 .

4.5 Testes à forma funcional da regressão

Testes à forma funcional da regressão são uma parte importante de qualquer análise de um modelo. Nesta secção vamos estudar alguns desses testes. Para uma leitura mais completa recomenda-se Thursby [134] e Thursby et. al [135].

RESET (Regression Specification Error Test)

Ramsey [116] desenvolveu um teste estatístico para a qualidade da especificação do modelo de regressão linear gaussiano acrescentando um conjunto de regressores Z , tal que o modelo aumentado é

$$Y = X^\top \beta + Z^\top \gamma. \quad (4.44)$$

Se Z fosse conhecido, então o teste reduz-se a um teste F para $H_0 : \gamma = 0$. O ponto crucial é a escolha das variáveis Z . Esta escolha depende da verdadeira forma funcional do modelo sobre a hipótese alternativa, que normalmente é desconhecida. Esta dificuldade pode ser ultrapassada usando potências das variáveis do modelo original como descrito em [116]. Outra forma de proceder consiste em usar momentos de ordem mais elevada de $\hat{Y} = X^\top \hat{\beta}$. Para efectuar este teste é usual proceder da seguinte forma:

1. aplicar o método dos mínimos quadrados a (4.44) para encontrar \hat{Y}
2. aplicar o método dos mínimos quadrados a

$$Y = \hat{Y} + \gamma_2 \hat{Y}^2 + \dots + \gamma_m \hat{Y}^m \quad (4.45)$$

Sobre a hipótese nula, tem-se

$$\frac{\frac{R_1^2 - R_0^2}{m}}{\frac{1 - R_1^2}{n - (m+k) - 1}} \sim F_{(k-1; n-k-m)} \quad (4.46)$$

onde R_0^2 e R_1^2 são o valor de R^2 respectivamente da primeira e segunda regressão.

Thursby [134] desenvolveu um procedimento para fazer a distinção entre erros de especificação da forma funcional do modelo e heterocedasticidade. Esse procedimento consiste em usar simultaneamente o teste Reset e o teste de Goldfeld-Quandt (ver parágrafo (4.7.3) mais abaixo). Thursby deduz a independência dos dois testes e a utilidade de ambos. Conclui que, se o teste Reset é insensível à heterocedasticidade, então, independentemente do resultado do teste de Goldfeld-Quandt, uma estatística de teste Reset significativa implica que o modelo não está devidamente especificado. No caso contrário, ou seja, se tivermos um valor da estatística significativo para o teste de Goldfeld-Quandt e não significativo para o teste Reset, estamos na presença de heterocedasticidade. Estatísticas com valores não significativos em ambos os testes implica que não existe nenhum problema de heterocedasticidade.

Teste rainbow

A ideia básica por trás do teste Rainbow (ver [141]) é que, mesmo quando a verdadeira relação entre a variável resposta e as variáveis explicativas é não-linear, um bom ajustamento linear pode ser obtido sobre subconjuntos da amostra. Por conseguinte, o teste rejeita a hipótese nula de linearidade sempre que o valor de R^2 no ajuste global é significativamente inferior ao valor de R^2 sobre um, adequadamente seleccionado, subconjunto dos dados.

Seja $\epsilon^\top \epsilon$ a soma do quadrado dos resíduos pelo método dos mínimos quadrados para as n observações e seja $\bar{\epsilon}^\top \bar{\epsilon}$ a soma do quadrado dos resíduos obtidos pelo método dos mínimos quadrados a partir meio das observações ($n/2$ observações). Então, sobre a hipótese nula H_0 , o teste da hipótese

$$\begin{aligned} H_0 : \sigma^2 &= \mathbb{E}(\epsilon^\top \epsilon / (n - k)) &= \mathbb{E}(\bar{\epsilon}^\top \bar{\epsilon} / (n/2 - k)) \\ H_1 : \sigma^2 &< \mathbb{E}(\epsilon^\top \epsilon / (n - k)) &< \mathbb{E}(\bar{\epsilon}^\top \bar{\epsilon} / (n/2 - k)) \end{aligned}$$

segue a seguinte distribuição

$$\frac{\frac{\epsilon^\top \epsilon - \bar{\epsilon}^\top \bar{\epsilon}}{n/2}}{\frac{\bar{\epsilon}^\top \bar{\epsilon}}{n/2 - k}} \sim F_{(n/2; n/2 - k)} \quad (4.47)$$

A questão crucial no teste Rainbow é a escolha adequada do subconjunto a usar na construção da estatística. Isto afecta a potência do teste e, naturalmente, a distribuição da estatística de teste sob a hipótese nula. Utts [141] recomenda pontos próximos de \bar{X} , para a construção da amostra, uma vez que um incorrecto ajuste linear, em geral, não irá ser tão flagrante nas proximidades de \bar{X} como seria em regiões mais afastadas. A proximidade a \bar{X} é medido pela magnitude dos elementos diagonais correspondentes da matriz H , matriz alavanca na regressão. Pontos próximos são aqueles com baixa alavancagem h_i . O tamanho ideal do subconjunto a testar depende da hipótese alternativa. Utts recomenda que se use cerca de metade das observações, a fim de obter alguma robustez em relação a outliers.

4.6 Transformações para a homocedasticidade

No chamado método de regressão linear, ver por exemplo [40], a transformação da variável dependente está suficientemente generalizada para lidar com a heterocedasticidade, dependendo da distribuição dos dados. Se a variância pode ser representada como uma forma conhecida dependente da média, o investigador pode resolver a questão da heterocedasticidade usando uma transformação de variável conveniente. Por exemplo, se o investigador suspeitar que a variância é proporcional ao quadrado da média, então é possível aplicar a transformação logarítmica à variável da resposta. Por outro lado, se a variância parece ser directamente proporcional à média então, usando a transformação da raiz quadrada da variável de resposta, obteremos um modelo homocedástico.

Uma transformação da variável resposta, no entanto, deve ser usada com discrição, pois se os dados são lineares e heterocedásticos, se for aplicada uma transformação desajustada podemos destruir a linearidade e assim vir a reduzir a precisão das previsões obtidas por modelos de regressão que assumem linearidade.

Outro desafio que surge está naturalmente na selecção da transformação adequada. Um dos métodos mais comuns para identificar o tipo de heterocedasticidade que depende da média consistem em analisar o gráfico dos quadrados dos resíduos contra os valores ajustados no modelo de regressão [44, 138]. No entanto, apesar deste procedimento ser simples, identificar uma transformação apresenta algumas dificuldades, isto porque os dados podem ter características tais como outliers ou serem não lineares. Um caso ilustrativo está representado na Figura 4.5, em que uma simples transformação logarítmica evidencia uma forma simples de eliminar a heterocedasticidade que resultou bastante bem nestes dados. A forma como medimos os dados ou observações em diversos casos não reflecte as relações entre eles e, tal como Tuckey afirma em [137], aprender a transformar os dados é uma arte subtil da análise de dados. Veremos que no problema que nos interessa investigar essa questão é colocada. Colocam-se as seguintes questões:

- Como escolher essa transformação?
- Haverá alguma transformação dos dados que facilite a análise de regressão?
- Como decidir uma transformação é a mais adequada aos dados em estudo?

Uma transformação revela-se proveitosa dividem-se em três tipos de situações:

1. Variável resposta ter uma distribuição não normal conhecida. Para solucionar esta problema é escolhida uma transformação tal que a variável transformada tenha uma distribuição o mais próximo possível da distribuição normal;
2. o valor esperado da resposta, $\mathbb{E}(Y)$, está relacionado com as variáveis explicativas X_i através de uma função não linear nos parâmetros;
3. a distribuição dos erros e a forma funcional da regressão não conhecidas de forma precisa. E a que apresenta mais dificuldades em solucionar.

Como idealmente se pretende uma transformação que forneça um modelo facilmente interpretável e que faça sentido cientificamente, com variância constante e erros aproximadamente normais, relembramos procedimentos habitualmente usados para resolver este tipo de problemas, para uma descrição mais detalhada consultar [39].

Transformação de potências

É a família de transformações mais conhecida:

$$\Psi(U, \lambda) = U^\lambda \quad (4.48)$$

O parâmetro λ pode assumir vários valores, por regra entre -2 e 2 , mas usualmente restringe-se ao intervalo -1 e 1 , sendo que os mais facilmente identificados são:

- $\lambda = 1/2$ para a raiz quadrada;
- $\lambda = 1/3$ para a raiz cúbica;
- $\lambda = 0$ para o logaritmo natural;
- $\lambda = -1$ para a transformação inversa;
- $\lambda = 1$, em que não se faz qualquer transformação.

Fox et. al, [64], descreve um processo em que se podem usar métodos empíricos para se encontrar um valor optimizado de λ , usando o gráfico de cada um das variáveis explicativas contra os valores ajustados do modelo, e visualizando uma linha que reflecta a relação entre esses valores e a variável explicativa.

Naturalmente em muitos destes casos, a variável a transformar tem de ser estritamente positiva.

Método de Box-Cox

Quando a distribuição normal não se adequa aos dados, muitas vezes é útil aplicar a transformação de Box-Cox para obtermos a normalidade, ver [147]. Considerando x_1, \dots, x_n os dados originais, a transformação de Box-Cox [21], recorre à função de verosimilhança para estimar o valor de λ , ver Figura 4.6.

Esta transformação é dada por:

$$y_i^{(\lambda)} = \begin{cases} \frac{y_i^\lambda - 1}{\lambda} & \text{se } \lambda \neq 0, \\ \ln(y_i) & \text{se } \lambda = 0, \end{cases} \quad (4.49)$$

Regra de Mosteller e Tukey

Se uma relação é não linear, mas no entanto é monótona e a curvatura da função de regressão é relativamente constante, Mosteller e Tuckey, em [105] propõem uma regra ilustrada na figura 4.7, que pode guiar-nos na escolha de transformações para a linearização.

4.7 Pré-Testes

Pelo que já foi dito e pela consulta da literatura que existe um consenso generalizado segundo o qual é necessário tomar medidas correctivas na presença de dados heterocedásticos. Não existe, no entanto, nenhum teste estatístico para decidir da presença ou não de uma relação heterocedástica que seja comumente aceite. Ao longo dos anos têm sido desenvolvidos vários testes dos quais se podem citar os testes de *White*, *Goldfeld-Quandt*, *Glejser*, *Theil*, *Breuch-Pagan* e *Szroeter*, entre muitos outros, (ver [39, 45, 69]).

A maior parte destes testes são não construtivos, ou seja, são usados para determinar se estamos ou não na presença de heterocedasticidade, mas não revelam nada sob a forma funcional dessa heterocedasticidade. Testes construtivos, por outro lado, requerem a especificação e estimação de uma estrutura paramétrica da forma funcional da heterocedasticidade são testes do tipo Wald ou rácio de verosimilhança que permitem esta modelação da heterocedasticidade.

As formas de heterocedasticidade mais usadas são a aditiva e a multiplicativa.

- Aditiva tem a seguinte forma

$$\sigma_i^2 = Z_i^\top \gamma, \quad \text{para } i = 1, \dots, n.$$

- A heterocedasticidade multiplicativa investigada pela primeira vez por Park [114] e mais tarde aprofundada por Harvey [83] tem a seguinte forma:

$$\sigma_i^2 = \exp(Z_i^\top \gamma) \quad \text{para } i = 1, \dots, n.$$

Nas próximas sub-seções vamos descrever alguns destes testes, e comentar sobre a sua potência e introduziremos uma técnica de Monte Carlo para minimizar o problema de erros de Tipo I que estes testes apresentam (ver [58]).

4.7.1 Teste de White

Seja $\hat{\epsilon} = (\hat{\epsilon}_1, \dots, \hat{\epsilon}_n)$ o vector dos resíduos calculados usando o método dos mínimos quadrados e seja

$$\hat{\sigma}^2 = \frac{1}{n-k} \hat{\epsilon}^\top \hat{\epsilon}$$

o estimador da sua variância assumindo a homocedasticidade. Quando a hipótese de homocedasticidade é verdadeira temos que

$$\hat{V} = \frac{1}{n} \sum_{i=1}^n \hat{\epsilon}_i^2 x_i x_i^\top$$

e

$$\bar{V} = \frac{1}{n} \sum_{i=1}^n \hat{\sigma}^2 x_i x_i^\top$$

são estimadores consistentes da mesma matriz de covariância. No entanto, caso a hipótese nula seja falsa, estas tendem a divergir. Ou seja, construindo uma estatística baseada neste

facto podemos avaliar da veracidade da hipótese nula. O que White, [153] propõe é que, na versão mais simples, se considere a estatística nR^2 , onde R^2 é calculado da regressão $\hat{\epsilon}^2$, onde $\hat{\epsilon}$ são os resíduos da regressão original, nas variáveis $X^\top \otimes X^\top$, onde são eliminadas as variáveis redundantes. Sobre a hipótese nula nR^2 tem uma distribuição qui-quadrado com graus de liberdade igual ao número de variáveis desta última regressão. Ou seja, se não existirem redundâncias em $X^\top \otimes X^\top$, então o número de variáveis será $k(k+1)/2 - 1$. White em [153] defini as condições suficientes para que estes resultados sejam aplicáveis.

Tal como White afirma, este teste também é um teste de má especificação da forma funcional da regressão original, pois, por exemplo, seja $x_i \in X$ uma variável, se a estatística de White mostrar evidências que devemos recusar a hipótese nula e o investigador não suspeitar de evidências de heterocedasticidade então o teste é conclusivo de que falta uma variável na forma funcional da regressão, por exemplo, x_i^2 .

4.7.2 Teste do coeficiente de correlação de Spearman

Yin et. al [156] sugere que se use o coeficiente de Spearman, para criar uma estatística para avaliar se a variância é constante.

O coeficiente de correlação de Spearman é uma medida não paramétrica de dependência estatística entre duas variáveis que, ao contrário do coeficiente de correlação de Pearson, não é afectado pela hipótese da distribuição das variáveis em estudo não serem normais. Para uma amostra de dimensão n e variáveis X e Y , os ranks de ambas as variáveis, x_i e y_i , são determinados e calcula-se a sua diferença $d_i = x_i - y_i$. O coeficiente de correlação de Spearman é

$$\rho_s = 1 - \left(6 \sum_{i=1}^n \frac{d_i^2}{n^3 - n} \right). \quad (4.50)$$

Yin et. al considera o seguinte procedimento:

- considera-se o valor absoluto dos resíduos da regressão original e uma das variáveis independentes X_j ;
- calcula-se o índice na ordenação desses mesmo resíduos e na ordenação da variável independente e determina a diferença entre esses índices:

$$d_i = \text{rank}(|\epsilon_i|) - \text{rank}(X_i);$$

- considere-se a hipótese nula

$$H_0 : \quad \rho = 0,$$

de que o coeficiente de correlação de Spearman é nulo.

- a estatística de teste é

$$T = \rho \left(\frac{n-2}{1-\rho^2} \right)^{1/2}$$

que tem uma distribuição t -Student com $n-2$ graus de liberdade;

- se o valor-p for superior ao nível de significância pretendido, não rejeitamos a hipótese nula de homocedasticidade, caso contrário, rejeitamos a hipótese nula.

Torna-se então necessário investigar o modelo quanto à presença de heterocedasticidade. Note-se que este resultado implica, caso se rejeite a hipótese nula, que existe uma correlação entre as variáveis em estudo, portanto os resíduos têm uma associação com a variável independente usada no teste.

4.7.3 Teste de Goldfeld-Quandt

Considere-se um conjunto de observações, X , de variáveis explicativas e Y uma variável resposta. Assuma-se que se pode dividir as observações em dois grupos tais que, sob a hipótese de homocedasticidade, os dois grupos apresentam a mesma variância, e que na hipótese alternativa as variâncias variam nos dois subgrupos, ou seja,

$$\begin{aligned} H_0 : \sigma_1^2 &= \sigma_2^2 \\ H_1 : \sigma_1^2 &\neq \sigma_2^2 \end{aligned} \quad (4.51)$$

onde σ_1^2 é a variância dos resíduos da regressão de X_1 em Y_1 com n_1 observações e σ_2^2 é a variância dos resíduos da regressão de X_2 em Y_2 com n_2 observações.

O que é sugerido por Goldfeld et. al [73], é que se ordene as observações de forma a separar as que apresentam maior variância das de menor variância. A estatística de teste requer dividir a amostra em dois grupos com n_1 e n_2 observações respectivamente. Para obter variâncias estatisticamente independentes estima-se a regressão para cada um dos subgrupos e toma-se a seguinte estatística

$$GQ = \frac{\epsilon_1^T \epsilon_1 / (n_1 - k)}{\epsilon_2^T \epsilon_2 / (n_2 - k)} \quad (4.52)$$

onde e_1 e e_2 são os vectores de resíduos de cada uma das regressões, com $n_1 \geq n_2$. Sob a hipótese nula tem-se

$$GQ \sim F_{(n_1-k, n_2-k)}.$$

Goldfeld et. al sugere que se omitam algumas observações no centro da amostra, no entanto, quantas mais observações se eliminar, menor serão os graus de liberdade em cada um dos grupos. Vários autores sugerem que se evite eliminar mais que 1/3 da amostra.

Quando a hipótese nula é verdadeira, a estatística GQ segue exactamente uma distribuição $F_{(n_1-k, n_2-k)}$, e a dimensão do teste é correcta; caso contrário, a distribuição é apenas aproximada e é recomendável usar um teste mais robusto como o teste de White.

Levene [98], sugere uma estatística semelhante à de Goldfeld-Quandt, que consiste em:

$$L_0 = \frac{\sum_{i=1}^g (\bar{z}_i - \bar{z})^2 / (g - 1)}{\sum_{i=1}^g \sum_{j=1}^{n_i} (\bar{z}_{ij} - \bar{z}_i)^2 / (\sum_{i=1}^g (n_i - 1))} \quad (4.53)$$

onde o conjunto dos resíduos, ϵ , é dividido em $g \leq n$ grupos com j observações, $1 \leq j \leq n_i$, sendo

$$z_{ij} = |\epsilon_{ij} - \bar{\epsilon}_i|, \quad (4.54)$$

$$\bar{z}_i = \left(\sum_{j=1}^{n_i} z_{ij} \right) / n_i \quad (4.55)$$

e

$$\bar{z} = \left(\sum_{i=1}^g \sum_{j=1}^{n_i} z_{ij} \right) / \left(\sum_{i=1}^g n_i \right). \quad (4.56)$$

Os valores críticos de L_0 são obtidos da distribuição F com $g - 1$ e $\sum_{i=1}^g (n_i - 1)$ graus de liberdade. Considerando a mediana $\tilde{\epsilon}_i$ de ϵ_i substituindo a média de ϵ_i pela sua mediana na equação (4.54) temos uma nova estatística:

$$L_1 = \frac{\sum_{i=1}^g (\tilde{z}_i - \bar{z})^2 / (g - 1)}{\sum_{i=1}^g \sum_{j=1}^{n_i} (\tilde{z}_{ij} - \tilde{z}_i)^2 / (\sum_{i=1}^g (n_i - 1))} \quad (4.57)$$

que esta a mesma distribuição da anterior.

Carapeto et. al, em [34], propõem um teste semelhante ao teste de Goldfeld-Quandt, onde as observações são ordenadas pela ordem decrescente da variável X_j que se considera ser responsável pela heterocedasticidade, e divididas em dois subgrupos, um com m observações e outro com $n - m$ observações.

Quer-se testar as hipóteses

$$\begin{aligned} H0 : \quad \sigma_1^2 &\leq \sigma_2^2 \\ H1 : \quad \sigma_1^2 &> \sigma_2^2 \end{aligned} \quad (4.58)$$

e a estatística teste é definida por

$$q = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} = \frac{\hat{\epsilon}^\top I_1 \hat{\epsilon}}{\hat{\epsilon}^\top I_2 \hat{\epsilon}} = \frac{\hat{\epsilon}^\top M^\top I_1 M \hat{\epsilon}}{\hat{\epsilon}^\top M^\top I_2 M \hat{\epsilon}} \quad (4.59)$$

onde I_1 e I_2 são semelhantes à matriz identidade, com dimensão $n \times n$, em que alguns dos elementos da diagonal são zero, mais especificamente, no caso de I_1 , os elementos da sua diagonal não nulos são os m elementos de X_j que têm maior variância, enquanto que I_2 representa os restantes. Usando a relação $\hat{\epsilon} = M\epsilon$ pois a matriz de projecção é $M = I - X(X^\top X)^{-1}X^\top$.

Dado um resultado particular q_0 para a estatística, podemos calcular $P[q > q_0]$ e assim obter um valor p para o teste

$$P \left[\frac{\epsilon^\top M^\top I_1 M \epsilon}{\epsilon^\top M^\top I_2 M \epsilon} > q_0 \right] = P[\epsilon^\top M^\top (I_1 - q_0 I_2) M \epsilon > 0]. \quad (4.60)$$

A distribuição das formas quadráticas $M^\top I_1 M$ e $M^\top I_2 M$, matrizes simétricas semi-positivas quadráticas, são calculadas usando o algoritmo de Imhof [88], que é uma avaliação numérica da inversão do integral de Fourier, sob a hipótese de que os resíduos tem distribuição normal com média zero. O valor- p é representado por

$$P[\epsilon^\top M^\top (I_1 - q_0 I_2) M \epsilon > 0] = \frac{1}{2} + \frac{1}{\pi} + \int_0^\infty \frac{\sin(\theta(t))}{t\rho(t)} dt \quad (4.61)$$

onde $\theta(t) = \frac{1}{2} \sum_{j=1}^n \tan^{-1}(\lambda_j t)$, $\rho(t) = \prod_{j=1}^n (1 + \lambda_j^2 t^2)^{1/4}$ e λ_j são os valores próprios de $M^\top(I_1 - q_0 I_2)M$.

4.7.4 Teste de Breusch-Pagan

O teste de Breuch-Pagan é descrito em [25]. Este teste é construído considerando a seguinte hipótese

$$\sigma_i^2 = \sigma^2 f(\gamma_0 + z_i^\top \gamma) \quad (4.62)$$

onde z_i é um vector de variáveis independentes. O modelo é homocedástico se $\gamma = 0$. A estatística teste é calculada fazendo a regressão de $\epsilon_i^2/(\epsilon^\top \epsilon/n)$ em z_i e obtendo a soma dos quadrados explicada pela regressão. Ou seja, é dado por:

$$BP = \frac{1}{2} (g^\top Z (Z^\top Z)^{-1} Z^\top g) \quad (4.63)$$

onde $g_i = \epsilon_i^2/(\epsilon^\top \epsilon/n) - 1$. Sob a hipótese nula BP tem uma distribuição qui-quadrado com número de graus de liberdade igual ao número de variáveis do vector z_i .

Koenker, [93] argumenta que este teste não é robusto quando a distribuição dos resíduos não é normal e sugere o seguinte estimador para a variância de ϵ_i^2 ,

$$V = \frac{1}{n} \sum_{i=1}^n \left(\epsilon_i^2 - \frac{\epsilon^\top \epsilon}{n} \right)^2. \quad (4.64)$$

De facto, tendo ϵ_i uma distribuição que não a normal, a variância deste não é necessariamente $2\sigma^2$. Seja então $u = (\epsilon_1^2, \dots, \epsilon_n^2)$ e $\mathbf{1}$ o vector coluna, de dimensão $n \times 1$, composto por uns. Então

$$\bar{u} = \frac{\epsilon^\top \epsilon}{n}$$

e (4.63) pode-se

$$\bar{BP} = \frac{1}{V} (u - \bar{u} \mathbf{1})^\top Z (Z^\top Z)^{-1} Z^\top (u - \bar{u} \mathbf{1}) \quad (4.65)$$

Sob a hipótese nula, esta estatística tem a mesma distribuição que a primeira. Note-se que a estatística proposta por Koenker é uma versão studentizada da versão de Breuch-Pagan.

4.7.5 Teste de Szroeter

Szroeter [131] introduziu uma classe de estimadores da forma

$$\bar{h} = \frac{\sum_{i \in M} h_i \epsilon_i^2}{\sum_{i \in M} \epsilon_i^2} \quad (4.66)$$

onde M é um subconjunto de $\{1, 2, \dots, n\}$, e os h_i formam uma sequência crescente de escalares não estocásticos dos pesos. Szroeter sugere várias maneiras de obter valores para os pesos h_i . Uma dessas maneiras baseia nos resíduos do método dos mínimos quadrados.

Propõe os seguintes estimadores:

$$SKH = \left(2 \sum_{i=1}^n \left(1 - \cos \left(\frac{\pi i}{i+1} \right) \right) \hat{\epsilon}_i^2 \right) / \left(\sum_{i=1}^n \hat{\epsilon}_i^2 \right) \quad (4.67)$$

$$S_N = \left(\frac{6n}{n^2 - 1} \right)^{1/2} \left(\frac{\sum_{i=1}^n i \hat{\epsilon}_i^2}{\sum_{i=1}^n \hat{\epsilon}_i^2} - \frac{n+1}{2} \right) \quad (4.68)$$

$$S_F = \left(\sum_{i=n_1+n_2+1}^n \hat{\epsilon}_i^2 \right) / \left(\sum_{i=1}^{n_1} \hat{\epsilon}_i^2 \right) \quad (4.69)$$

Sob as condições da hipótese nula, S_N segue, assintoticamente, uma distribuição $\mathcal{N}(0, 1)$. Se os resíduos, ϵ_i seguirem uma distribuição

$$\epsilon \stackrel{iid}{\sim} \mathcal{N}(0, 1)$$

então os pontos críticos de SKH podem ser encontrados usando o método de Imhof, [88]. Szroeter sugere os seguintes limites para a estatística SKH : h_L^* para o limite inferior e h_U^* para o limite superior, sendo h_L^* e h_U^* os limites do teste de Durbin-Watson, [59], aplicando a um conjunto de $n + 1$ observações e k variáveis explicativas. Rejeita-se a hipótese nula se $SKH > 4 - h_L^*$ e não se rejeita caso $SKH < 4 - h_U^*$, considerando-se inconclusivo caso contrário.

A estatística S_F pode ser interpretada como uma variante da estatística GQ (4.52), onde os resíduos de regressões dos subgrupos foram substituídos pelos da regressão original baseados em todas as observações, em (4.52), tal que $\epsilon_1^T \epsilon_1$ é substituído por

$$\sum_{i=n_1+n_2+1}^n \hat{\epsilon}_i^2$$

e $\epsilon_2^T \epsilon_2$ é substituído por

$$\sum_{i=1}^n \hat{\epsilon}_i^2.$$

Harrison et. al [81] sugerem uma estatística semelhante baseado na razão entre a soma dos quadrados de um subconjunto de resíduos e a soma dos quadrados dos resíduos do conjunto total:

$$HM = \left(\sum_{i=1}^{n_1} \hat{\epsilon}_i^2 \right) / \left(\sum_{i=1}^n \hat{\epsilon}_i^2 \right), \quad (4.70)$$

onde $n_1 = \lfloor n/2 \rfloor$.

Harrison et. al [81] propõem os seguintes limites para esta estatística:

$$b_L^* = (1 + b(n - n_1, n - k, n - k))^{-1}$$

como limite inferior;

$$b_U^* = (1 + b(n - n_1 - k, n_1, n - k))^{-1}$$

como limite superior, onde $b = (v_1, v_2, v_3) = (v_1/v_3)F_\alpha(v_1, v_3)$ e $F_\alpha(v_1, v_2)$ é o valor crítico para o nível α da distribuição $F(v_1, v_2)$. A hipótese nula, homocedasticidade dos resíduos, é rejeitada se $HM < b_L^*$ e não rejeitada se $HM > b_U^*$; caso contrário o teste é inconclusivo. Harrison sugere aproximações para a estatística HM , ver [80], mas estas, infelizmente, oferecem uma redução de custos computacionais desprezável.

4.7.6 Teste de Monte Carlo

Assumindo que

$$\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^\top \quad (4.71)$$

tem uma distribuição contínua condicional em X então, em situações em que a amostra é finita e se suspeita da baixa potência do teste que está a ser usado para aferir da homocedasticidade, de entre os testes anteriormente descritos, podemos recorrer a um procedimento de Monte Carlo como o descrito por Dufour et. al [58].

Suponhamos que se pretende verificar se uma estatística, digamos $S_0 = S(Y, X)$, cuja distribuição nula, dado X , é livre de parâmetros. Seja $G(x)$ a função de sobrevivência associada a S_0 sobre a hipótese H_0 , isto é

$$G : \mathbb{N} \longrightarrow [0, 1]$$

é a função tal que

$$G(x) = P_{H_0}[S_0 \geq x] \quad \forall x$$

onde P_{H_0} é a medida de probabilidade nas condições da hipótese nula. Supondo que a distribuição de S_0 é contínua, então temos que $G(x) = 1 - F(x)$ onde

$$F(x) = P_{H_0}[S_0 \leq x]$$

é a função de distribuição de S_0 sobre H_0 .

Sem perda de generalidade, suponhamos que o argumento é para um ponto crítico na cauda à direita: H_0 é rejeitada para um nível α quando $S_0 > c(\alpha)$, onde $c(\alpha)$ é o valor do nível crítico tal que $G(c(\alpha))$, ou equivalentemente, quando o valor- p associado à estatística é menor ou igual a α .

Suponhamos que é possível gerar N replicações, independentes e identicamente distribuídas, do vector ε nas condições de (4.71). Este procedimento permite obter N simulações e N realizações independentes da estatística teste, S_1, \dots, S_N .

Descrevemos de seguida o algoritmo para o teste de Monte Carlo:

1. Gerar N replicações independentes e identicamente distribuídas do vector ε
2. fazer $Y = X^\top \beta + \varepsilon^{(i)}$ para $i = 1, \dots, N$
3. ajustar o modelo $Y = X^\top \hat{\beta}^{(i)}$
4. determinar a estatística S_i

5. criar o vector A , com comprimento N
6. se $\hat{p}_i(S_i) \leq \alpha$ fazer $A[i] = 1$; se $\hat{p}_i(S_i) > \alpha$ fazer $A[i] = 0$
7. calcular $\hat{G}_N = \frac{\sum_{i=1}^N A[i] + 1}{N + 1}$
8. se $\hat{G}_N \leq \alpha$ rejeita-se a hipótese nula de homocedasticidade para um nível α ; caso contrário não se rejeita a hipótese nula.

A região crítica associada a este procedimento de Monte Carlo é

$$\hat{p}_N(S_0) \leq \alpha$$

$$\begin{aligned}\hat{p}_N(x) &= \frac{N\hat{G}_N(x) + 1}{N + 1} \\ \hat{G}_N(x) &= \frac{\sum_{i=1}^N \mathbb{1}_{[0, \infty[}(S_i - x)}{N}\end{aligned}$$

sendo $\mathbb{1}$ a função indicatriz, como é habitual:

$$\mathbb{1}_A(x) = \begin{cases} 1, & \text{se } x \in A \\ 0, & \text{se } x \notin A \end{cases}.$$

Então, desde que a distribuição de S_0 induzida por P_{H_0} na condição de H_0 seja contínua, temos

$$P_{H_0}[p_N(S_0) \leq \alpha] = \frac{I[\alpha(N + 1)]}{N + 1}, \quad \text{para } 0 \leq \alpha \leq 1. \quad (4.72)$$

Note-se que a adição de 1 no denominador e numerador em (4.72) é necessário para esta se verificar. Em particular, se N for escolhido tal que $\alpha(N + 1)$ seja um inteiro, temos que $P_{H_0}[p_N(S_0) \leq \alpha] = \alpha$. Logo a região crítica (4.72) tem a mesma dimensão que $G(x) \leq \alpha$. Este procedimento assim obtido é teoricamente exacto, independentemente do número N de replicações usado.

Uma implicação da hipótese (4.71) é que a normalidade não tem qualquer interferência na validade do teste, logo é possível usar qualquer distribuição para ϵ como, por exemplo, distribuições de cauda pesada.

Como exemplo, dessa necessidade, considere-se o seguinte modelo

$$Y = 1.5X + \epsilon, \quad (4.73)$$

onde $X \sim \mathcal{N}(0, 1)$, é um vector, 25×1 , e $\epsilon \sim \mathcal{N}(0, X^2)$, é um vector, 25×1 . Simulando este modelo para $N = 100$ e usando o teste de Breusch-Pagan obtemos a Figura 4.8, de onde concluímos que, por exemplo, o teste em questão no que toca a este modelo não é muito eficiente a detectar a heterocedasticidade. E pela Figura 4.8, podemos também concluir que mesmo aumentando o tamanho da amostra a eficiência deste teste não melhora. O que demonstra a necessidade de usar o método de Monte Carlo neste tipo de situações.

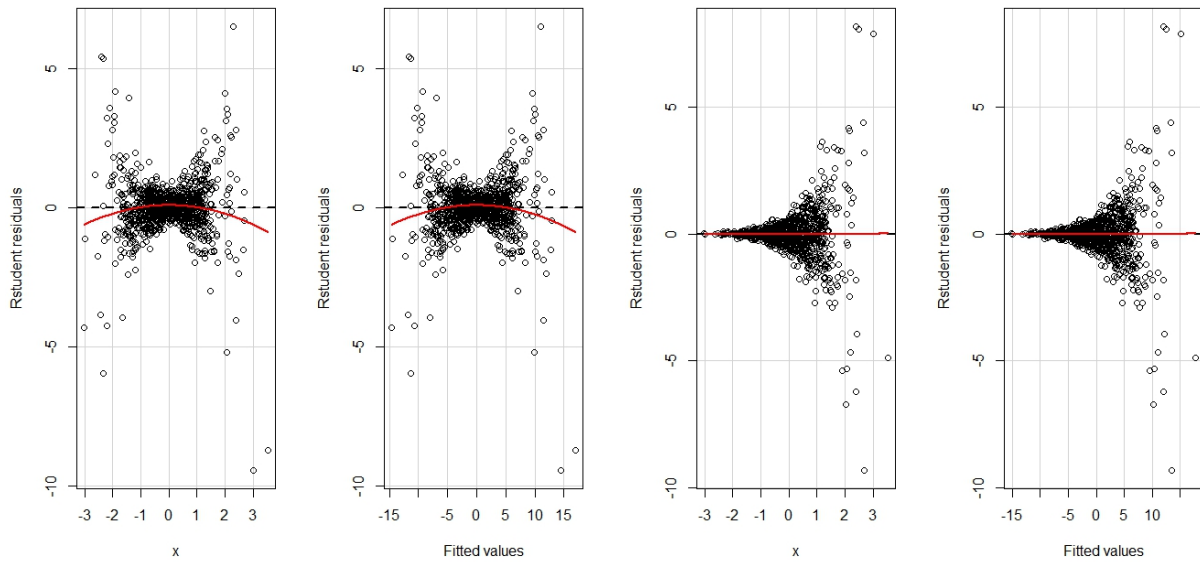
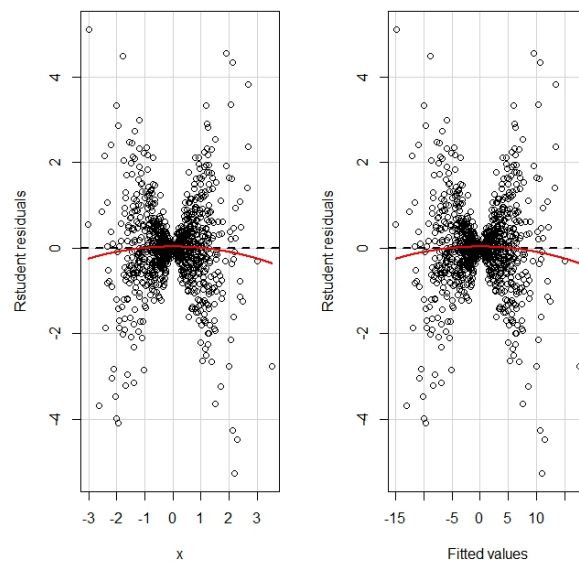
(a) $w_i = \exp(|z_i|)$ (b) $w_i = \exp(z_i)$ (c) $w_i = |z_i|$

Figura 4.2: Gráfico dos resíduos estudentizados para aferir da forma da heterocedasticidade, onde os nomes dos eixos são respectivamente na horizontal a variável independente $z = x$ e $z = y\hat{\beta}$, e na vertical os valores dos resíduos estudentizados.

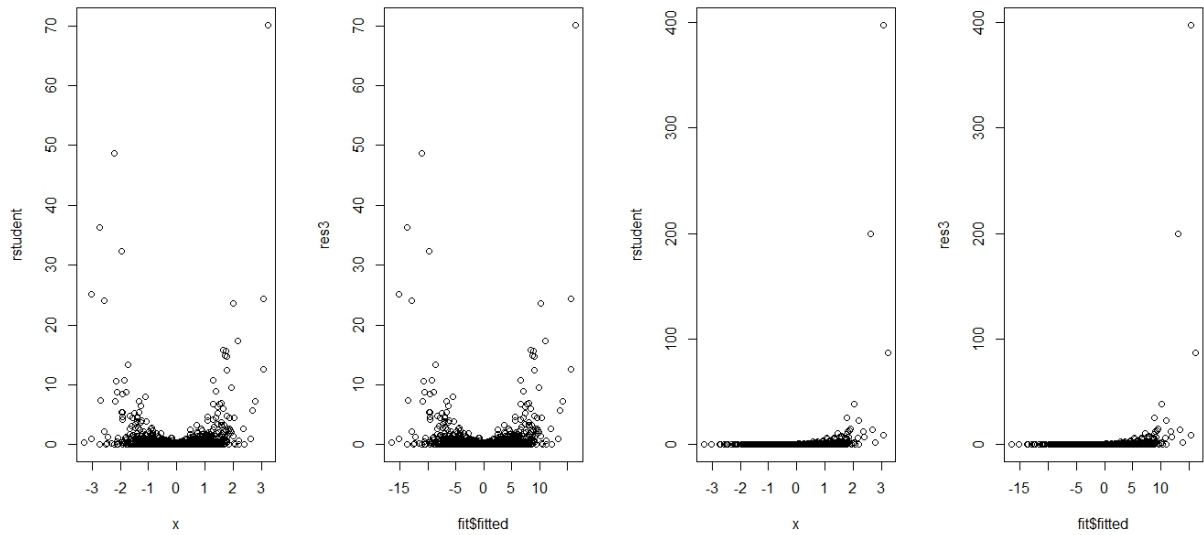
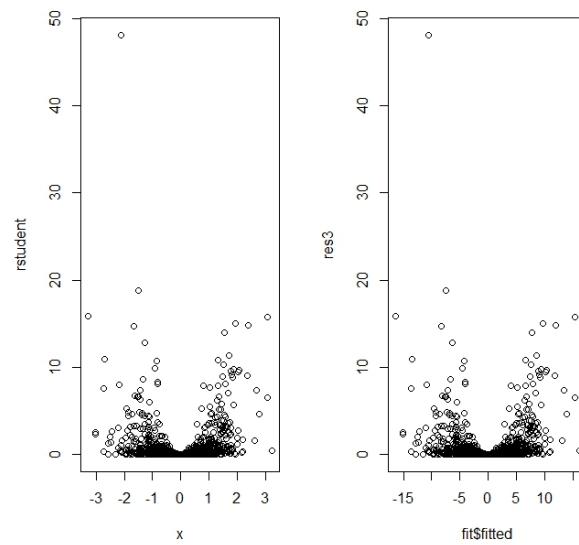
(a) $w_i = \exp(|z_i|)$ (b) $w_i = \exp(z_i)$ (c) $w_i = |z_i|$

Figura 4.3: Gráfico dos resíduos estudentizados para aferir da forma da heterocedasticidade, onde os eixos são respectivamente na horizontal a variável independente $z = x$ e $z = y\hat{\beta}$, e na vertical os valores dos resíduos estudentizados quadráticos.

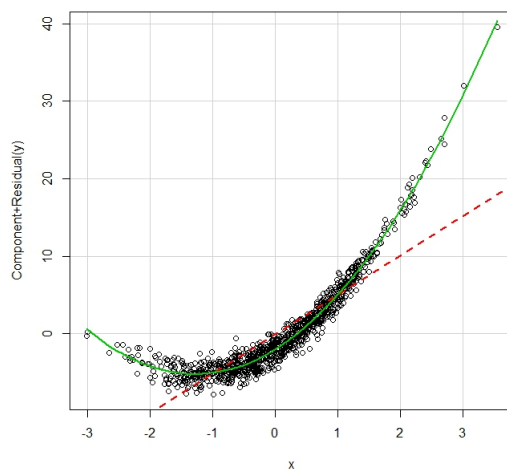


Figura 4.4: Gráfico parcial de resíduos do modelo 4.43.

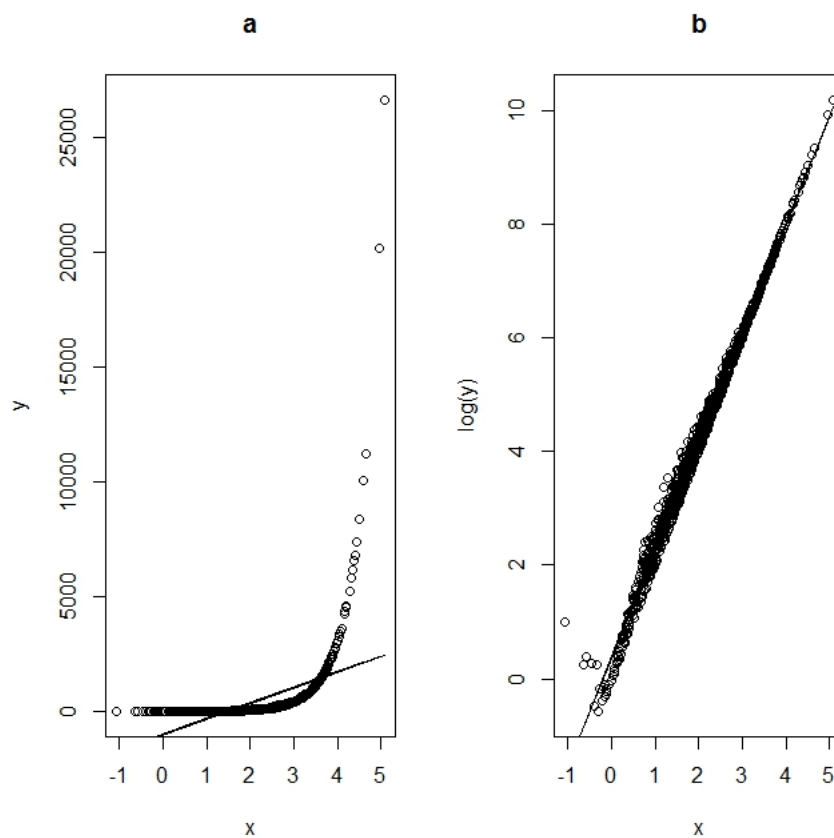


Figura 4.5: (a) Gráfico de dispersão em que a variância da resposta é proporcional ao quadrado da média, (b) Gráfico de dispersão dos mesmos dados depois de aplicada a transformação logarítmica

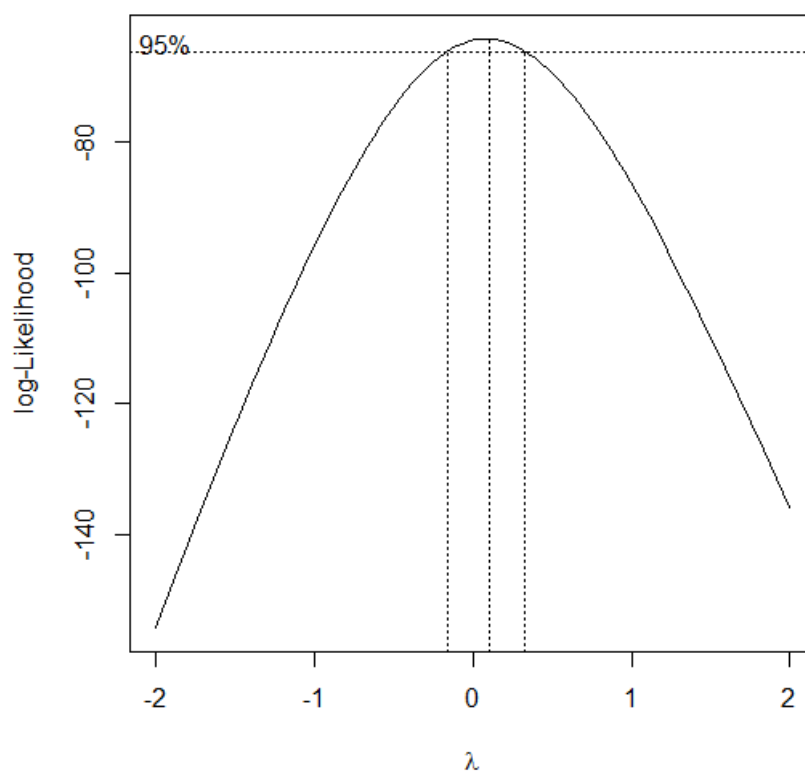


Figura 4.6: Transformação de Box-Cox

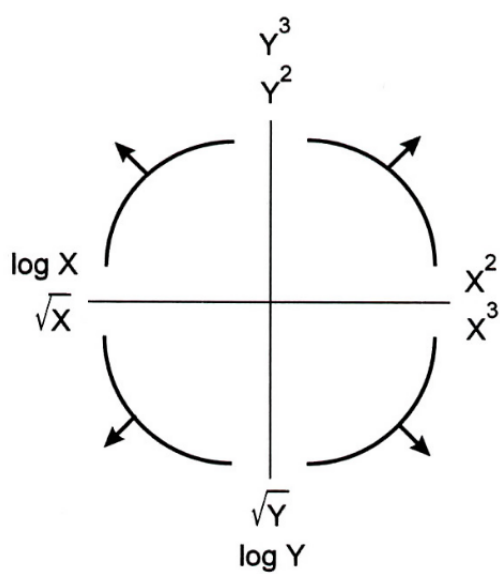


Figura 4.7: Regra de Mosteller e Tukey

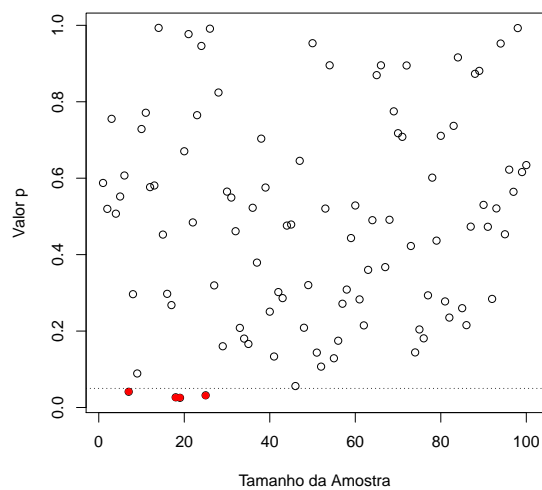


Figura 4.8: Diagrama de dispersão do valor-p do teste de Breusch-Pagan consoante o tamanho da amostra

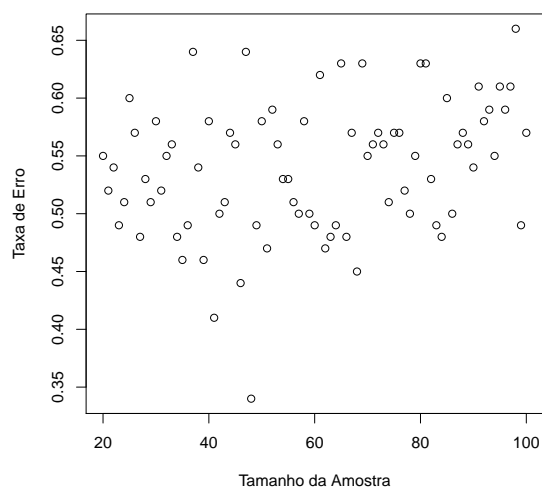


Figura 4.9: Diagrama de dispersão da taxa de erro para o teste de Breusch-Pagan consoante o tamanho da amostra

4.8 Modelação da heterocedasticidade

Uma vez apresentados diferentes testes a que podemos recorrer para decidir sobre a presença ou não de heterocedasticidade, passamos agora a expor diversas formas de modelar a heterocedasticidade.

4.8.1 Variância constante entre subgrupos

Seja

$$\begin{aligned} Y_i &= X_i^\top \beta + \epsilon_i \quad \text{para } i = 1, \dots, n \\ \mathbb{E}(\epsilon_i) &= 0 \end{aligned} \quad (4.74)$$

Suponhamos que as n observações estão divididas em m grupos cada, um deles com n_m observações; o vector β é o mesmo em todos os grupos mas, para cada m , tem-se

$$\text{var}(\epsilon_{in}) = \sigma_m^2, \quad \text{para } i = 1, \dots, n_m \quad (4.75)$$

Então, fazendo $\sigma_m^2 = s_m^2$, o estimador dos mínimos quadrados ponderados vem

$$\hat{\beta} = \left(\sum_{i=1}^M \left(\frac{1}{s_m^2} \right) X_m^\top X_m \right)^{-1} \left(\sum_{i=1}^M \left(\frac{1}{s_m^2} \right) X_m^\top Y_m \right) \quad (4.76)$$

Como $X_m^\top Y_m = X_m^\top X_m \hat{\beta}_m$ onde $\hat{\beta}_m$ é o estimador pelo método dos mínimos quadrados do m subconjunto de observações, então

$$\begin{aligned} \hat{\beta} &= \left(\sum_{i=1}^M \left(\frac{1}{s_m^2} \right) X_m^\top X_m \right)^{-1} \left(\sum_{i=1}^M \left(\frac{1}{s_m^2} \right) X_m^\top X_m \hat{\beta}_m \right) \\ &= \left(\sum_{i=1}^M V_m \right)^{-1} \left(\sum_{i=1}^M V_m \hat{\beta}_m \right) \\ &= \sum_{i=1}^M W_m \hat{\beta}_m \end{aligned} \quad (4.77)$$

onde

$$W_m = \left(\sum_{i=1}^M (\text{var}(\beta_m))^{-1} \right)^{-1} (\text{var}(\beta_m))^{-1}. \quad (4.78)$$

Ou seja, o estimador com a menor matriz de covariância recebe o maior peso e vice-versa. Note-se que, se o número de observações for menor do que o número de regressores, os estimadores $\hat{\beta}_m$ não podem ser calculados e este resultado (4.77) não tem qualquer utilidade. No entanto, não é isso que se pretende. O que se pretende é estimar os pesos. Para tal, como os estimadores pelo método dos mínimos quadrados, apesar de enviesados, são consistentes,

podemos usar os resíduos encontrados por este método para calcular

$$\hat{\sigma}_m^2 = \frac{\hat{\epsilon}^\top \hat{\epsilon}}{n_m}$$

Esta expressão fornece o estimador necessário para determinar as variâncias de cada subgrupo. Com este estimador podemos obter o estimador da matriz de covariância assintótica.

4.8.2 Desvio padrão de Y é uma função linear de uma ou mais variáveis independentes

Considere-se o seguinte modelo:

$$\begin{aligned} Y &= X\beta + \epsilon \\ \mathbb{E}(\epsilon) &= 0 \\ \mathbb{E}(\epsilon^2) &= \sigma^2 = (Z^\top \gamma)^2 \\ \mathbb{E}(\epsilon_i \epsilon_j) &= 0 \quad \text{para } i \neq j \end{aligned} \tag{4.79}$$

onde γ é um vector de parâmetros desconhecidos de dimensão $p \times 1$ com $0 < p \leq k$, e Z é um vector de variáveis, que podem ou não ser idênticas a uma variável da matriz X ou serem uma função de uma variável de X . O estimador de β pelo método dos mínimos quadrados generalizado é

$$\begin{aligned} \hat{\beta} &= (X^\top \Omega^{-1} X)^{-1} X^\top \Omega^{-1} Y \\ &= \left(\sum_{i=0}^n (z_i^\top \gamma)^{-2} x_i x_i^\top \right)^{-1} \sum_{i=0}^n (z_i^\top \gamma)^{-2} x_i y_i \end{aligned} \tag{4.80}$$

Quando γ é substituído por um estimador, $\hat{\gamma}$, em (4.80), ficamos com um estimador pelo método dos mínimos quadrados ponderados.

Por hipótese, os erros $\sigma_1^{-1}\epsilon_1, \sigma_2^{-1}\epsilon_2, \dots, \sigma_n^{-1}\epsilon_n$ são independentes e identicamente distribuídos com média zero e variância 1, logo

$$\mathbb{E}(\sigma_i^{-1}|\epsilon_i|) = c \quad \text{para } 1 \leq i \leq n$$

onde c é uma constante, independente de i e depende da distribuição de $\sigma_i^{-1}|e_i|$. Daqui se deduz que

$$|\hat{\epsilon}_i| = (Z^\delta)^\top \gamma + v \tag{4.81}$$

onde $\hat{\epsilon}_i$ são os resíduos do modelo (4.79) e v são os novos erros, tais que

$$v = |\hat{\epsilon}| - \mathbb{E}(|\epsilon|). \tag{4.82}$$

Aplicando o método dos mínimos quadrados a (4.81) temos que um estimador de $c\gamma$ é

$$c\hat{\gamma} = (Z^\top Z)^{-1} Z^\top |\hat{\epsilon}|. \tag{4.83}$$

Usando agora o método dos mínimos quadrados ponderados, temos que um estimador robusto de β é

$$\hat{\hat{\beta}} = \left(\sum_{i=1}^n (z_i^\top \gamma)^{-2} x_i x_i^\top \right)^{-1} \sum_{i=1}^n (z_i^\top \gamma)^{-2} x_i y_i \quad (4.84)$$

Note-se que o estimador $\hat{\gamma}$, em geral, não possui as propriedades requeridas para a aplicação do método dos mínimos quadrados, pois v é heterocedástico, e auto-correlacionado e com média não nula, mas se estiver nas condições apropriadas, ver [83], $|\hat{\epsilon}|$ converge em distribuição para $|\epsilon|$. O estimador $\hat{\gamma}$ será um estimador consistente de γ e o estimador de β terá as mesmas propriedade assintóticas do estimador dos mínimos quadrados generalizado.

Assimptoticamente não é possível melhorar o estimador $\hat{\hat{\beta}}$, mas, encontrando um melhor estimador de γ encontraremos um melhor estimador de β . Para tal, note-se que, se $|\hat{\epsilon}|$ converge para $|\epsilon|$, então assintoticamente v será independentemente distribuído com média zero e variância dada por

$$\mathbb{E}(v^2) = \mathbb{E}(|\epsilon^2|) - (\mathbb{E}(|\epsilon|))^2 = \sigma^2(1 - c)^2 = (Z^\top \gamma)^2(1 - c)^2 \quad (4.85)$$

Aplicando o método dos mínimos quadrados ponderados a γ temos

$$c\hat{\gamma} = \left(Z^\top \hat{\Omega}^{-1} Z \right)^{-1} \left(Z^\top \hat{\Omega}^{-1} |\hat{\epsilon}| \right). \quad (4.86)$$

O segundo estimador é descrito por Harvey em [83], e o primeiro por Glejser, em [68], com o propósito de testar a homocedasticidade.

O teste que Glejser, [68] parte da hipótese de homocedasticidade, ou seja,

$$\begin{aligned} H_0 : & \sigma_i = \sigma \\ H_1 : & \exists_i : \sigma_i \neq 0 \quad \text{para } i = 1, \dots, n \end{aligned} \quad (4.87)$$

Para tal, considere-se $\gamma^* = 0$, onde $\gamma^* = (\gamma_2, \dots, \gamma_k)^\top$ e $\gamma' = (\gamma_1, \gamma^{*\top})$. Então, supondo a hipótese nula verdadeira, temos que a covariância de $\hat{\gamma}$ é:

$$\Sigma = \left(\frac{1 - c^2}{c^2} \right) \gamma_1^2 (Z^\top Z)^{-1}, \quad (4.88)$$

onde γ_1^2 é estimado pelos resíduos do modelo (4.81) e $\gamma_1^2 = \epsilon^\top \epsilon$. Logo se $\hat{\gamma}$ e $\hat{\gamma}^*$ forem variáveis com distribuição normal, temos que as estatísticas teste são

$$G_{\hat{\gamma}} = \left(\frac{c^2}{1 - c^2} \right) \frac{\hat{\gamma}^{*\top} D^{-1} \hat{\gamma}^*}{\hat{\gamma}_1^2} \quad (4.89)$$

e

$$G_{\hat{\gamma}} = \left(\frac{c^2}{1 - c^2} \right) \frac{\hat{\gamma}^{*\top} D^{-1} \hat{\gamma}^*}{\hat{\gamma}_1^2}, \quad (4.90)$$

que seguem uma distribuição qui-quadrado com $k - 1$ graus de liberdade, \mathcal{X}_{k-1}^2 , e $D = (Z^\top Z)^{-1}$. Se os erros são normalmente distribuídos, então

$$c = \left(\frac{2}{\pi} \right)^{1/2}.$$

A hipótese nula é rejeitada se as estatísticas $G_{\hat{\gamma}}$ ou $G_{\hat{\gamma}}$ são superiores a $\mathcal{X}_{(k-1, \alpha)}^2$. A estatística $G_{\hat{\gamma}}$ é computacionalmente mais conveniente que $G_{\hat{\gamma}}$, pois $c^2 \hat{\gamma}^{*\top} D^{-1} \hat{\gamma}^*$ é igual à soma dos quadrados da regressão de $|\hat{\epsilon}|$ em Z .

Glejser, (ver [68]) deduz um processo mais simples para testar a homocedasticidade. Este teste consiste em avaliar a significância dos parâmetros $\hat{\gamma}^{*\top}$ na regressão (4.81), caso qualquer um destes parâmetros seja significativo rejeita-se a hipótese nula de homocedasticidade.

Considere-se com por exemplo o modelo (4.79) em que $X \sim \mathcal{N}(0, 1)$ é um vector, 1000×1 , e $\epsilon \sim \mathcal{N}(0, X)$ é um vector, 1000×1 , daqui se obtêm as Figuras 4.11 e 4.10 que serve de exemplo para comparação com possíveis modelos com heterocedasticidade da forma do modelo (4.79).

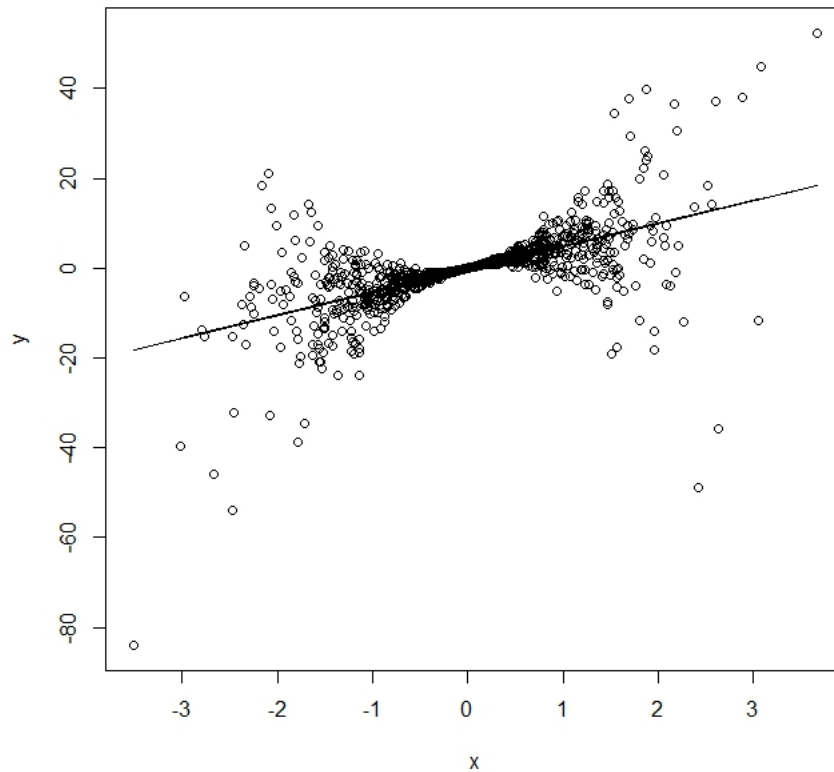


Figura 4.10: Gráfico de dispersão entre X e Y para o modelo (4.79)

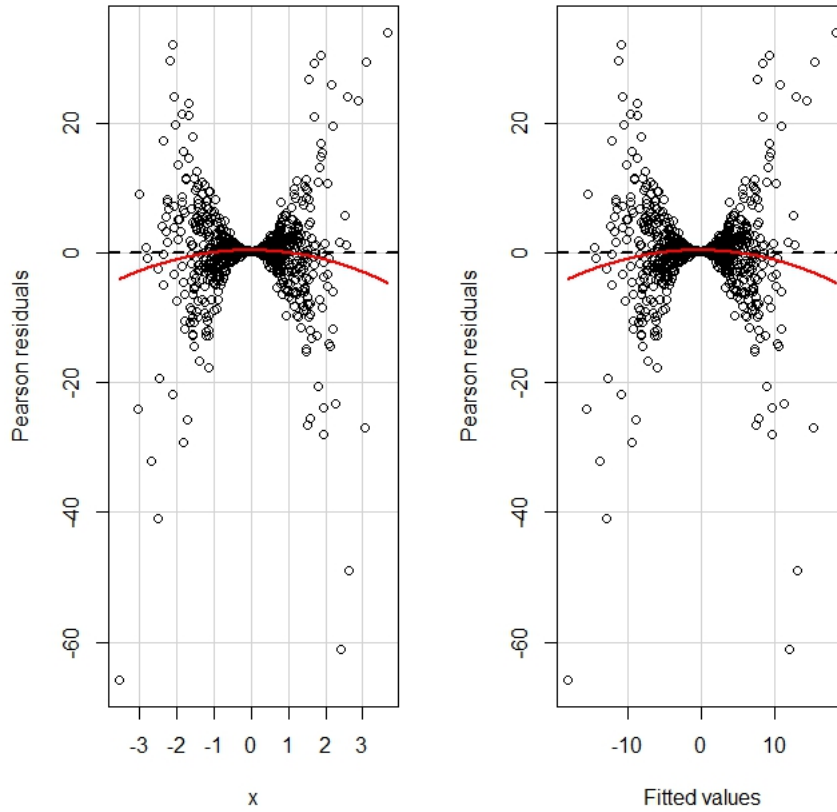


Figura 4.11: Gráfico dos resíduos para o modelo (4.79)

4.8.3 Variância de Y é uma função linear de uma ou mais variáveis independentes

Um outro modelo heterocedástico proposto por vários investigadores, tais como Amemiya [9], Harvey [83] e Goldfeld et. al [71] consiste em que σ^2 seja uma função linear de uma das variáveis explicativas, ou seja, quando temos o seguinte modelo:

$$\begin{aligned}
 Y &= X\beta + \epsilon \\
 \mathbb{E}(\epsilon) &= 0 \\
 \mathbb{E}(\epsilon^2) &= \sigma^2 = Z^\top \gamma \\
 \mathbb{E}(\epsilon_i \epsilon_j) &= 0 \quad \text{para } i \neq j
 \end{aligned} \tag{4.91}$$

onde γ e Z são como na secção 4.8.2.

O estimador β pelo método dos mínimos quadrados ponderados é

$$\hat{\beta} = (X^\top \Omega^{-1} X)^{-1} X^\top \Omega^{-1} Y \tag{4.92}$$

tem média β e covariância $(X^\top \Omega^{-1} X)^{-1}$.

Para o modelo (4.91) podemos escrever que

$$\hat{\epsilon}^2 = Z^\top \gamma + v \quad (4.93)$$

onde $v = \hat{\epsilon}^2 - \mathbb{E}(\epsilon^2)$ e $\hat{\epsilon}$ é o vector de resíduos do modelo (4.91). Então, como assinalado por Goldfeld [73], temos que:

$$\hat{\gamma} = (Z^\top Z)^{-1} Z^\top \dot{\hat{\epsilon}} \quad (4.94)$$

$\dot{\hat{\epsilon}} = (\hat{\epsilon}_1^2, \dots, \hat{\epsilon}_n^2)$. No entanto, v tem média nula, é heterocedástico e auto-correlacionado (ver [83]), logo $\hat{\gamma}$ é um estimador enviesado e ineficiente de γ . De facto note-se que se tem

$$\mathbb{E}(\dot{\hat{\epsilon}}) = \dot{M} Z \gamma,$$

onde \dot{M} é a matriz

$$\dot{M} = I_n - X(X^\top X)^{-1} X^\top,$$

em que cada um dos elementos da sua diagonal representa uma diferença de quadrados e $\mathbb{E}(\hat{\gamma}) = (Z^\top Z)^{-1} Z^\top \dot{M} Z \gamma$, o que implica que podemos ter os seguintes estimadores

$$\hat{\gamma} = (Z^\top \dot{M} \dot{M} Z)^{-1} Z^\top \dot{M} \dot{\hat{\epsilon}} \quad (4.95)$$

e

$$\hat{\hat{\gamma}} = (Z^\top \dot{M} Z)^{-1} Z^\top \dot{\hat{\epsilon}}, \quad (4.96)$$

onde (4.94) resulta de aplicar o método dos mínimos quadrados a

$$\dot{\hat{\epsilon}} = \dot{M} Z \gamma + w \quad (4.97)$$

de onde se pode concluir que $\mathbb{E}(w) = 0$, enquanto que (4.97) é um estimador Minque (ver [118]).

Considere-se como exemplo do modelo (4.91) a variável $X \sim \mathcal{N}(0, 1)$ é um vector, 1000×1 , e $\epsilon \sim \mathcal{N}(0, X)$ é um vector com dimensão 1000×1 . Daqui se obtêm as figuras 4.11 e 4.10 que serve de exemplo para comparação com possíveis modelos com heterocedasticidade da forma do modelo (4.91).

4.8.4 Variância de Y é proporcional a uma potência do valor esperado

Considere-se o seguinte modelo com erros heterocedásticos:

$$\begin{aligned} Y &= X\beta + \epsilon \\ \mathbb{E}(\epsilon) &= 0 \\ \mathbb{E}(\epsilon^2) &= \sigma^2 = \sigma^2(X^\top \beta)^2 \\ \mathbb{E}(\epsilon_i \epsilon_j) &= 0 \quad \text{para } i \neq j \end{aligned} \quad (4.98)$$

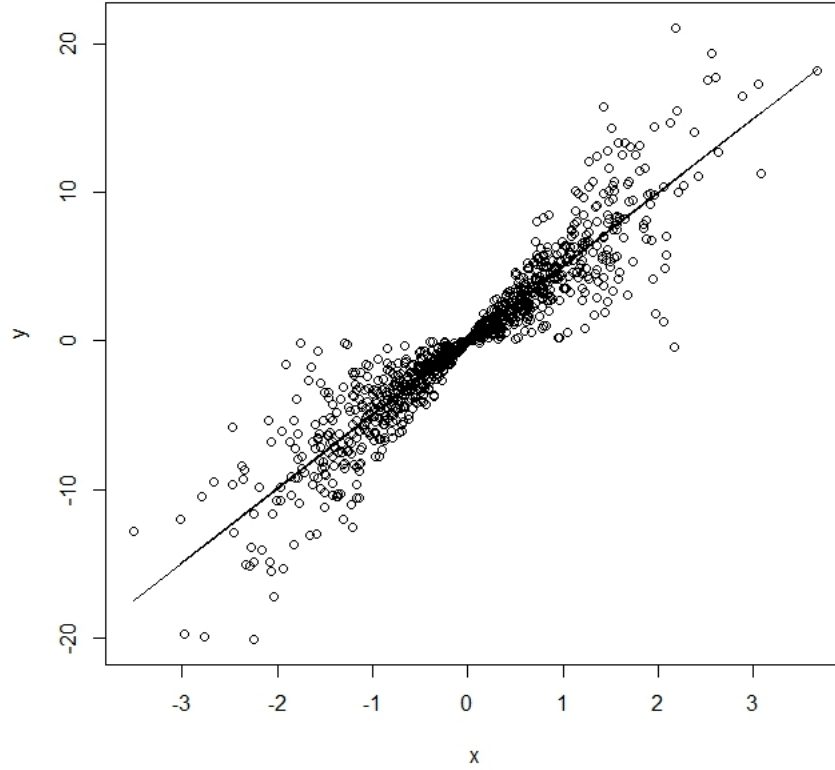


Figura 4.12: Gráfico de dispersão entre x e y para o modelo (4.91)

descrito por Theil [133] e por Amemiya [8]. A sua matriz de covariância é

$$\Omega = \sigma^2 \Psi = \sigma^2 \text{diag}((x_1^\top \beta)^2, \dots, (x_n^\top \beta)^2) \quad (4.99)$$

Este modelo pode ser visto como um caso particular de (4.79).

O estimador de β é dado por:

$$\hat{\beta} = (X^\top \Omega^{-1} X)^{-1} X^\top \Omega^{-1} Y.$$

Não é um estimador fiável, porque Ω depende de β e este é desconhecido mas, usando o método dos mínimos quadrados para obter um estimador usando $\hat{\beta} = (X^\top X)^{-1} X^\top Y$ e depois usando esta estimativa para obter uma estimativa pelo método dos mínimos quadrados ponderados através de

$$\hat{\beta} = (X^\top \hat{\Psi}^{-1} X)^{-1} X^\top \hat{\Psi}^{-1} Y, \quad (4.100)$$

nas condições descritas por Theil [133] e Amemiya [7], temos uma convergência para a normalidade, $\sqrt{n}(\hat{\beta} - \beta)$ tem uma distribuição normal com média 0 e matriz de covariância $\lim n((X^\top \hat{\Omega}^{-1} X)^{-1})$.

Considere-se o modelo (4.98) com $X \sim \mathcal{N}(0, 1)$ é um vector, 1000×1 , e $\epsilon \sim \mathcal{N}(0, X\beta)$, é um vector, 1000×1 , daqui se obtêm as figuras 4.11 e 4.10 que serve de exemplo para

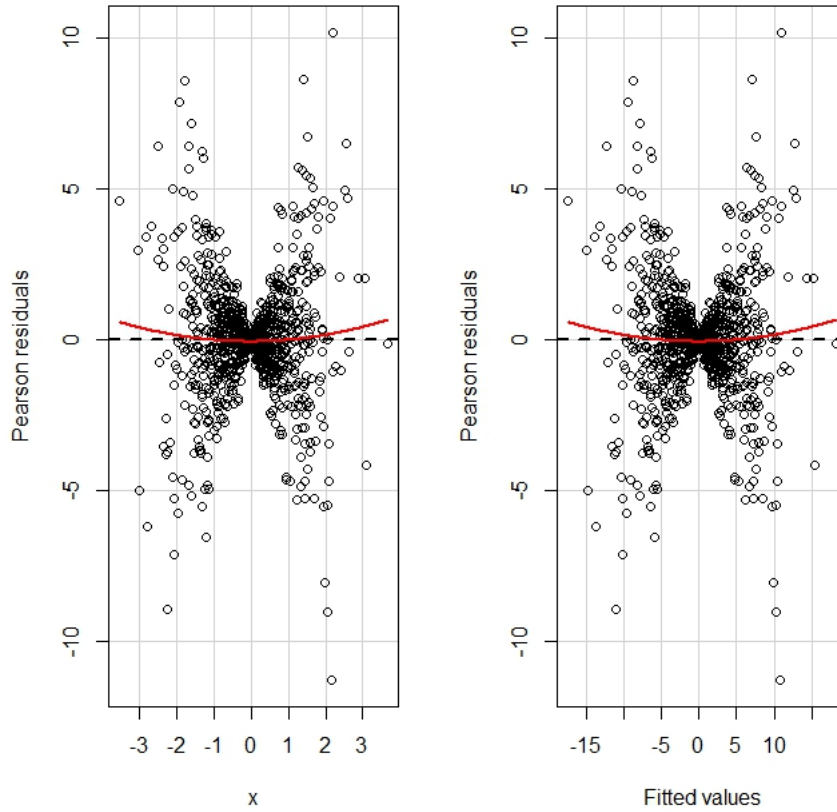


Figura 4.13: Gráfico dos resíduos para o modelo (4.91)

comparação com possíveis modelos com heterocedasticidade da forma do modelo (4.98).

4.8.5 Heterocedasticidade multiplicativa

Este modelo surge na literatura em 1973 com Harvey [83]. Harvey propõe os seguintes pressupostos para este modelo:

$$\begin{aligned}
 Y_i &= X_i^\top \beta + \epsilon_i \quad \text{para } i = 1, \dots, n \\
 \mathbb{E}(\epsilon_i) &= 0 \\
 \mathbb{E}(\epsilon_i^2) &= \sigma^2 = \exp(Z^\top \gamma) \\
 \mathbb{E}(\epsilon_i \epsilon_j) &= 0, \quad \text{para } i \neq j,
 \end{aligned} \tag{4.101}$$

onde Z é um vector nas condições dos modelos anteriores, n é naturalmente o número de observações e γ é um vector de parâmetros desconhecidos de dimensão $s \times 1$. Considere-se

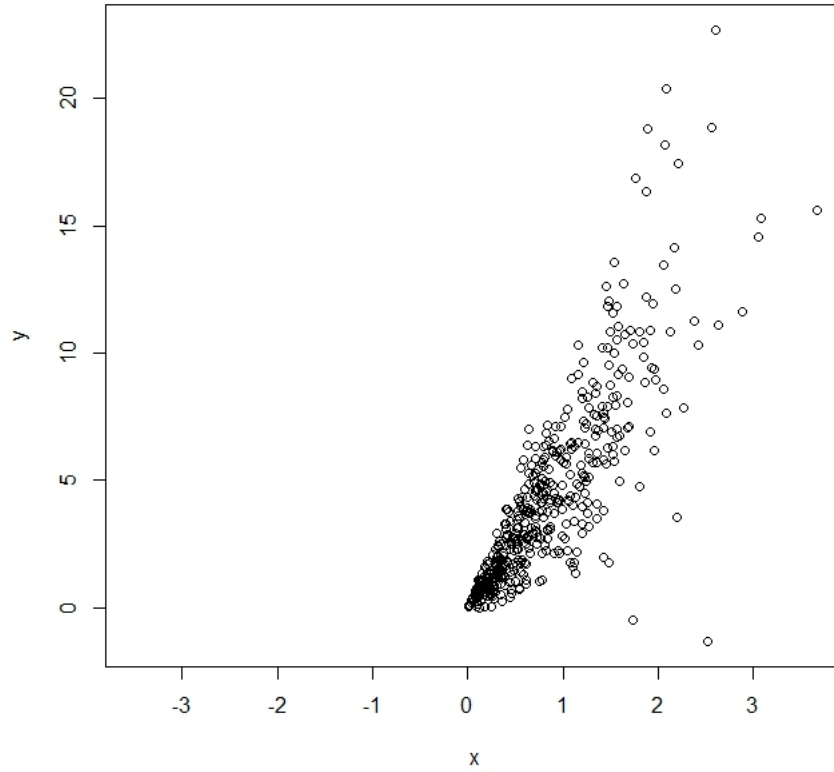


Figura 4.14: Gráfico de dispersão entre x e y para o modelo 4.98

o seguinte ilustrativo,

$$\begin{aligned} s &= 2 \\ \gamma^\top &= (\log k, \lambda) \\ Z_i^\top &= (1, \log X_{2i}). \end{aligned} \quad (4.102)$$

Então, temos que

$$\sigma_i^2 = kX_{2i}^\lambda,$$

o que significa que a variância é proporcional à potência de uma das variáveis explicativas, tal como descrito em [114, 94].

O estimador de β pelo método dos mínimos quadrados ponderados é obtido por:

$$\hat{\beta} = \left(\sum_{i=1}^n (\exp(-Z_i^\top \gamma) X_i X_i^\top) \right)^{-1} \sum_{i=1}^n (\exp(-Z_i^\top \gamma) X_i Y_i) \quad (4.103)$$

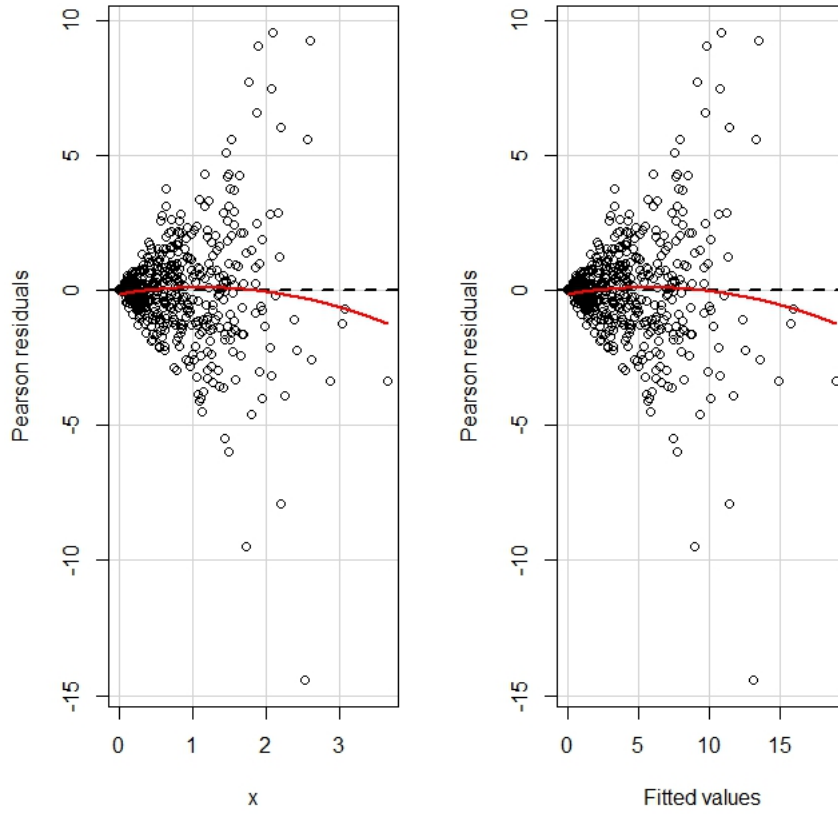


Figura 4.15: Gráfico dos resíduos para o modelo 4.98

Para obter o estimador $\hat{\beta}$, é necessário obter um estimador $\hat{\gamma}$ de γ . para isso note-se que, se fizermos

$$\log \hat{\epsilon}_i^2 = Z_i^\top \gamma + v_i, \quad (4.104)$$

onde

$$v_i = \log(\hat{\epsilon}_i^2 / \sigma_i^2),$$

aplicando o método dos mínimos quadrados a (4.104) temos

$$\hat{\gamma} = \left(\sum_{i=1}^n Z_i Z_i^\top \right)^{-1} \sum_{i=1}^n Z_i \log \hat{\epsilon}_i^2. \quad (4.105)$$

No entanto as perturbações v_i são heterocedásticos e auto-correlacionadas, com média não nula.

Admitindo que ϵ_i tem distribuição normal e se $\hat{\epsilon}_i$ converge em distribuição para ϵ_i , então assintoticamente, v_i é independente com média e variância dados por, (ver [83]):

$$\begin{aligned} \mathbb{E}(v_i) &= -1.2704 \\ \mathbb{E}(v_i^2) - (\mathbb{E}(v_i))^2 &= 4.9348 \end{aligned} \quad (4.106)$$

Isto implica que o primeiro elemento de $\hat{\gamma}$, $\hat{\gamma}_0$ é inconsistente. No entanto, a falta de consistência do estimador não tem consequências, pois na expressão do estimador

$$\hat{\beta} = \left(\sum_{i=1}^n (\exp(-Z_i^\top \gamma) X_i X_i^\top) \right) \sum_{i=1}^n (\exp(-Z_i^\top \gamma) X_i Y_i) \quad (4.107)$$

mudar $\hat{\gamma}_0$ muda apenas a constante de proporcionalidade. Se fizermos $\hat{\gamma}^\top = (\hat{\gamma}_1, \hat{\gamma}^{*\top})$, então $\hat{\gamma}^{*\top}$ é consistente e $\hat{\beta}$ é um estimador assintoticamente eficiente. Tem-se então que a matriz de covariância de $\hat{\gamma}$ é

$$\Sigma = 4.9348 (Z^\top Z)^{-1} \quad (4.108)$$

Daqui se deriva um teste de homocedasticidade, representada pela condição $\gamma^* = 0$. Considere-se a estatística

$$G_{\hat{\gamma}} = \left(\frac{c^2}{1 - c^2} \right) \frac{\hat{\gamma}^{*\top} D^{-1} \hat{\gamma}^*}{\hat{\gamma}_1^2} \quad (4.109)$$

que segue uma distribuição qui-quadrado com $s - 1$ graus de liberdade, sendo $D = (Z^\top Z)^{-1}$.

Park [114], deduziu um teste para este modelo de variância que passa por testar a significância do valor de γ_1 no modelo (4.104). Caso o teste t para $\gamma_1 = 0$ tenha um valor-p inferior a 0,05 então é rejeitada a hipótese nula de homocedasticidade e este modelo é admitido para a variância do modelo da média.

Considere-se o modelo (4.101) com $X \sim \mathcal{N}(0, 1)$ é um vector, 1000×1 , e $\epsilon \sim \mathcal{N}(0, \exp(X))$ é um vector, 1000×1 , daqui se obtêm as figuras 4.11 e 4.10 que serve de exemplo para comparação com possíveis modelos com heterocedasticidade da forma do modelo (4.101).

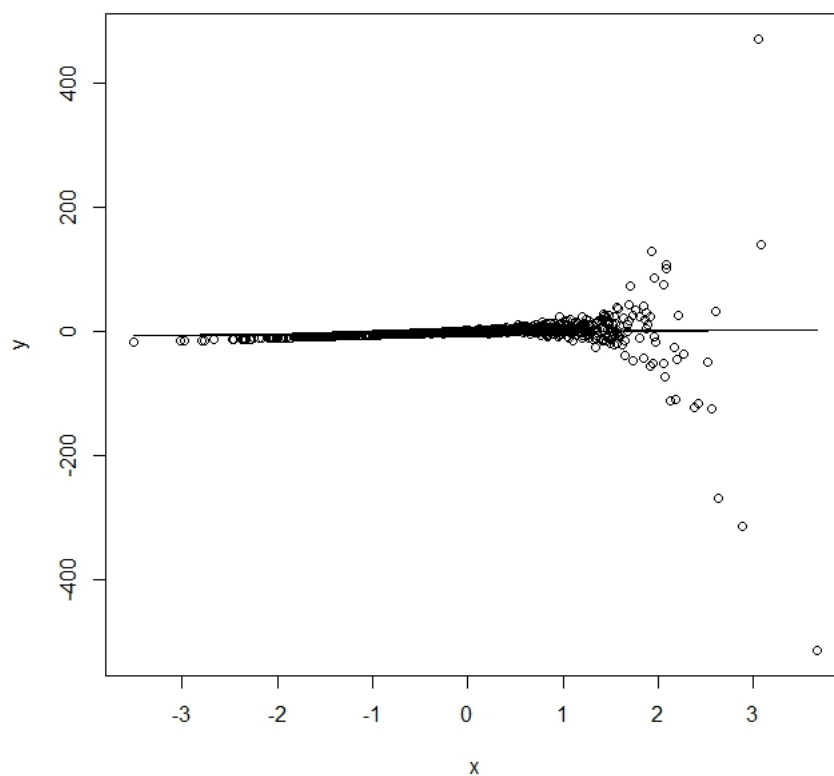


Figura 4.16: Gráfico de dispersão entre x e y para o modelo 4.101

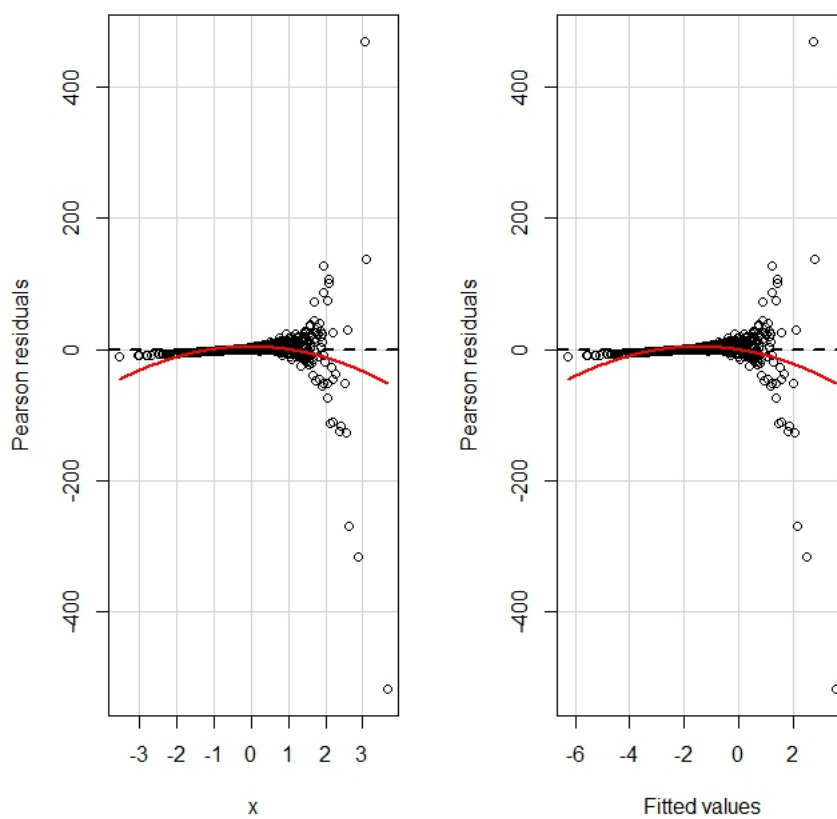


Figura 4.17: Gráfico dos resíduos para o modelo 4.101

4.8.6 Estimadores de White

Novamente, considere-se o modelo

$$Y = X\beta + \epsilon \quad (4.110)$$

Como já tivemos ocasião de referir anteriormente o estimador de β obtido pelo método dos mínimos quadrados é dado por $\hat{\beta} = (X^\top X)^{-1}X^\top Y$, logo

$$\text{var}(\hat{\beta}) = \Sigma = (X^\top X)^{-1}X^\top \Omega X (X^\top X)^{-1}. \quad (4.111)$$

Considere-se também a matriz "chapéu"

$$H = X(X^\top X)^{-1}X^\top, \quad (4.112)$$

onde os elementos da sua diagonal,

$$h_{ii} = x_i(X^\top X)^{-1}x_{ii}^\top,$$

são os valores de alavanca. Como

$$\epsilon = Y - X\hat{\beta} = (I_n - H)Y$$

temos, que, se considerarmos $\Omega = \sigma^2 I_n$ na expressão (4.111), está reduz-se a $\Sigma = \sigma^2 (X^\top X)^{-1}$. O que na prática se faz é estimar σ^2 usando o erro médio quadrático:

$$MSE = \frac{\sum_{i=1}^n \epsilon_i^2}{n - k} \quad (4.113)$$

do modelo de regressão (4.110), o que leva ao seguinte estimador:

$$\Sigma_{OLS} = MSE(X^\top X)^{-1}. \quad (4.114)$$

É desta ideia que White [153], entre outros, propõe: os estimadores HCCM. O primeiro estimador deste tipo, designado *HC0* surge com White:

$$\hat{\Sigma}_{HC0} = (X^\top X)^{-1}X^\top \text{diag}[\epsilon_i^2]X(X^\top X)^{-1}. \quad (4.115)$$

O estimador *HC0* é assintoticamente justificado, pois

$$\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{d} N(0, \Sigma), \quad (4.116)$$

mas, para pequenas amostras, o seu comportamento é enviesado, resultando em erros do Tipo I, (ver [52]).

Para obviar a esta falta de centricidade, Hinkley [84] propõe um outro estimador:

$$\hat{\Sigma}_{HC1} = \frac{n}{n - k} \hat{\Sigma}_{HC0}, \quad (4.117)$$

No entanto, este estimador não é recomendado pois apresenta os mesmos problemas com amostras finitas que *HC0*.

Outros investigadores propõem outras variantes deste estimador, baseados no método de "jackknife", que resultam de ajustar os valores da matriz chapéu, como por exemplo considerando:

$$\epsilon_i^* = \frac{\epsilon_i}{\sqrt{1 - h_{ii}}}$$

e construindo uma matriz diagonal com os valores de ϵ_i^* , de modo a obter-se o seguinte estimador:

$$\hat{\Sigma}_{HC2} = (X^\top X)^{-1} X^\top \text{diag} \left[\frac{\epsilon_i^2}{1 - h_{ii}} \right] X (X^\top X)^{-1}. \quad (4.118)$$

Sob a hipótese homocedasticidade é não enviesado, (ver [89]).

Se, em (4.118), substituirmos $\sqrt{(1 - h_{ii})}$ por $(1 - h_{ii})$ obtemos um estimador enviesado mas com melhor comportamento para erros do Tipo I, (consultar [102]),

$$\hat{\Sigma}_{HC3} = (X^\top X)^{-1} X^\top \text{diag} \left[\frac{\epsilon_i^2}{(1 - h_{ii})^2} \right] X (X^\top X)^{-1} \quad (4.119)$$

O estimador *HC3* apresenta bom comportamento em muitas situações, excepto quando os dados apresentam pontos de alavanca muito altos, (ver [152]), ou quando a distribuição dos erros é normalmente distribuída.

Cribari-Neto [52] propõem um estimador que tem em vista resolver o problema colocado por valores elevados de alavancagem. Para tal, consideram

$$\delta_i = \min \left\{ 4, \frac{nh_{ii}}{k} \right\},$$

ou seja, usa os valores de alavanca para descontar em (4.115), atribuindo maior peso a valores de alavancagem menores. O estimador proposto é dado por:

$$\hat{\Sigma}_{HC4} = (X^\top X)^{-1} X^\top \text{diag} \left[\frac{\epsilon_i^2}{(1 - h_{ii})^{\delta_i}} \right] X (X^\top X)^{-1} \quad (4.120)$$

Apesar de *HC4* apresentar um bom comportamento para os testes t , é um estimador mais enviesado do que *HC2* e *HC3*.

4.9 Comentários

Para um investigador que pretende estimar um modelo que admita a heterocedasticidade, as questões fundamentais são como testar o modelo quanto à presença de dados heterocedásticos e como testar o modelo heterocedástico escolhido. A conclusão a que se pode chegar é que não existe um melhor procedimento, cada processo tem os seus méritos e a análise dos resultados finais de cada modelo é que ditam o caminho a seguir.

Podemos adoptar uma variedade de testes que permitem avaliar da heterocedasticidade de um modelo, no entanto, como cada um dos testes tem características muito específicas, devem-se usar testes de carácter geral, tais como o teste de White ou de Breusch-Pagan. Note-se que, caso se ordene por ordem não decrescente, os resíduos do modelo que são

usados para estimar a variância deste, mas não se consiga distinguir, no caso de duas ou mais variáveis independentes, a direcção que a variância toma em relação a cada uma das variáveis independentes, o teste de Goldfeld-Quandt é inviável.

O investigador também deve ter em conta que, quando se testa a heterocedasticidade, mesmo um teste estatístico significativo, pode ser sintoma de uma má especificação do modelo, tal como uma variável não incluída no modelo ou a forma funcional da média incorrectamente especificada. Thursby [134] sugere uma estratégia para se fazer a discriminação entre cada uma das situações.

Se o investigador tiver à priori uma noção de qual a forma funcional da variância deve agir de acordo com essa convicção. No entanto, no caso do número de observações ser pequeno, é exigir muito aos dados atingir uma definição clara do modelo de variância. Surekha et. al [130], e concluíram que a eficiência de um estimador dos parâmetros de regressão obtido pelo método dos mínimos quadrados ponderados depende mais da escolha de estimador dos pesos e do tamanho da amostra do que da correcta especificação do modelo de variância. Os resultados por eles obtidos também implicam que estimadores baseados no método dos mínimos quadrados pesados não tem um bom comportamento para amostras de pequena dimensão, $n \leq 20$. Os mesmos investigadores mostraram ainda que a especificação

$$\sigma^2 = Z^\top \gamma$$

pode levar a estimadores fracos do parâmetro da regressão, caso sejam usadas variâncias negativas. Resumindo, a escolha de estimador para a variância não é importante, desde que não se use estimadores com propriedades assintóticas pobres. Na ausência de um conhecimento pormenorizado da forma da heterocedasticidade é recomendado recorrer aos estimadores de White. Para a estimação de um modelo específico a maioria dos estimadores recomendados exibem propriedades assintóticas idênticas, logo é razoável usar um qualquer estimador encontrado pelo método dos mínimos quadrados ponderados para o vector de parâmetros β que seja baseado nas propriedades de eficiência assintóticas dos parâmetros de estimação da variância.

No seguimento do trabalho anteriormente citado devido a Shureka et. al, num outro trabalho devido a Rilstone, [119], fica demonstrado que o uso de modelos semi-paramétricos em amostras de dimensão média, ($n = 50$) é a forma mais eficiente de modelar dados heterocedásticos, no caso de não se conhecer a forma da heterocedasticidade, sendo que o método dos núcleos se sobrepõem a qualquer outro método semi-paramétrico. Quanto à escolha da função de ponderação ou peso, não se demonstram grandes diferenças entre diversos tipos de especificação, mas o mesmo já não é verdade quanto ao parâmetro de suavização, onde se encontraram grandes diferenças. Também é recomendado comparar os resultados destes estimadores com os de White.

Capítulo 5

Aplicação na modelação da fluidez de pastas de betão

O que se pretende com este estudo é estabelecer uma relação entre o tempo de escoamento de 0,5 litro de pasta de cimento obtida através de um ensaio pelo Cone de Marsh e a água livre obtida pelo ensaio de centrifugação, para solucionar o problema deixado em aberto em [110], a heterocedasticidade do tempo de escoamento numa pasta de cimento no ensaio do cone de Marsh na relação com a água livre obtida pelo ensaio de centrifugação.

Foi elaborado um desenho factorial para o plano experimental visando estabelecer um modelo numérico entre o tempo de escoamento de uma pasta de cimento por um ensaio com o cone de Marsh (T_{flow}) e a água livre obtida pelo ensaio de centrifugação (W_{free}). Realizou-se um outro ensaio, o ensaio de espalhamento (D_{flow}), para controlo das observações.

Os dados usados neste estudo foram obtidos através de ensaios laboratoriais, tal como descrito no Capítulo 3, executados no Laboratório de Estruturas do Departamento de Engenharia Civil da Faculdade de Engenharia da Universidade do Porto. Os dados foram recolhidos usando formulário próprio disponível no Anexo A e tratados com o software estatístico R, versão 3.2.0. Os ficheiros com o código fonte em R encontram-se depositados em:

<https://www.dropbox.com/sh/sl8x1z3qw0s0w8p/AAADY4viH4RDUU6SXbAIx4n-a?dl=0>

O estudo permitirá analisar diferentes modelos, sugerindo dois modelos possíveis para a média das observações do tempo de escoamento usando o modelo de regressão linear gaussiano analisando várias formas de solucionar o problema da modelação da heterocedasticidade nos dados.

5.1 Breve discussão dos dados

Os dados obtidos são resumidos na Tabela 5.1 e na Figura 4.5.

Uma vez que Gomes, (ver [74]), recomenda que se use apenas as observações de T_{flow} abaixo de 55 s (pois observou que acima deste valor, o escoamento deixa de ser contínuo e há maiores diferenças entre os resultados de medições consecutivas). Procedemos à remoção das

	Wfree	Tflow	Dflow
Min. :	58.42	7.47	88.75
1st Qu.:	75.25	10.92	126.50
Median :	89.67	17.22	155.50
Mean :	92.68	23.93	150.40
3rd Qu.:	107.08	32.61	172.75
Max. :	143.75	78.52	212.00

Tabela 5.1: Sumário dos dados

correspondentes observações, temos que a actualização da anterior Tabela 5.1 é esplanada na Tabela 5.2.

	Wfree	Tflow	Dflow
Min.:	62.67	7.47	90.25
1st Qu.:	78.17	10.60	129.75
Median :	91.55	16.50	156.38
Mean :	94.23	21.21	152.52
3rd Qu.:	109.36	28.56	173.88
Max. :	143.75	54.66	212.00

Tabela 5.2: Sumário dos dados actualizada

Comecemos por analisar as possíveis relações que existam entre as variáveis estudadas. Para tal efectuem-se os habituais testes de correlação com os resultados apresentados na Tabela 5.3, para o coeficiente de correlação de Pearson, e na Tabela 5.4 para o coeficiente de correlação de Spearman. Daqui se conclui que existe uma forte relação entre as

	Wfree	Tflow	Dflow
Wfree	1	-0.844	0.337
Tflow	-0.844	1	-0.254
Dflow	0.337	-0.254	1

Tabela 5.3: Matriz de Correlação de Pearson

	Wfree	Tflow	Dflow
Wfree	1	-0.955	0.297
Tflow	-0.955	1	-0.308
Dflow	0.297	-0.308	1

Tabela 5.4: Matriz de Correlação de Spearman

variáveis *Wfree* e *Tflow*, pois tanto o coeficiente de Pearson, ($T=-0.844$), como o coeficiente

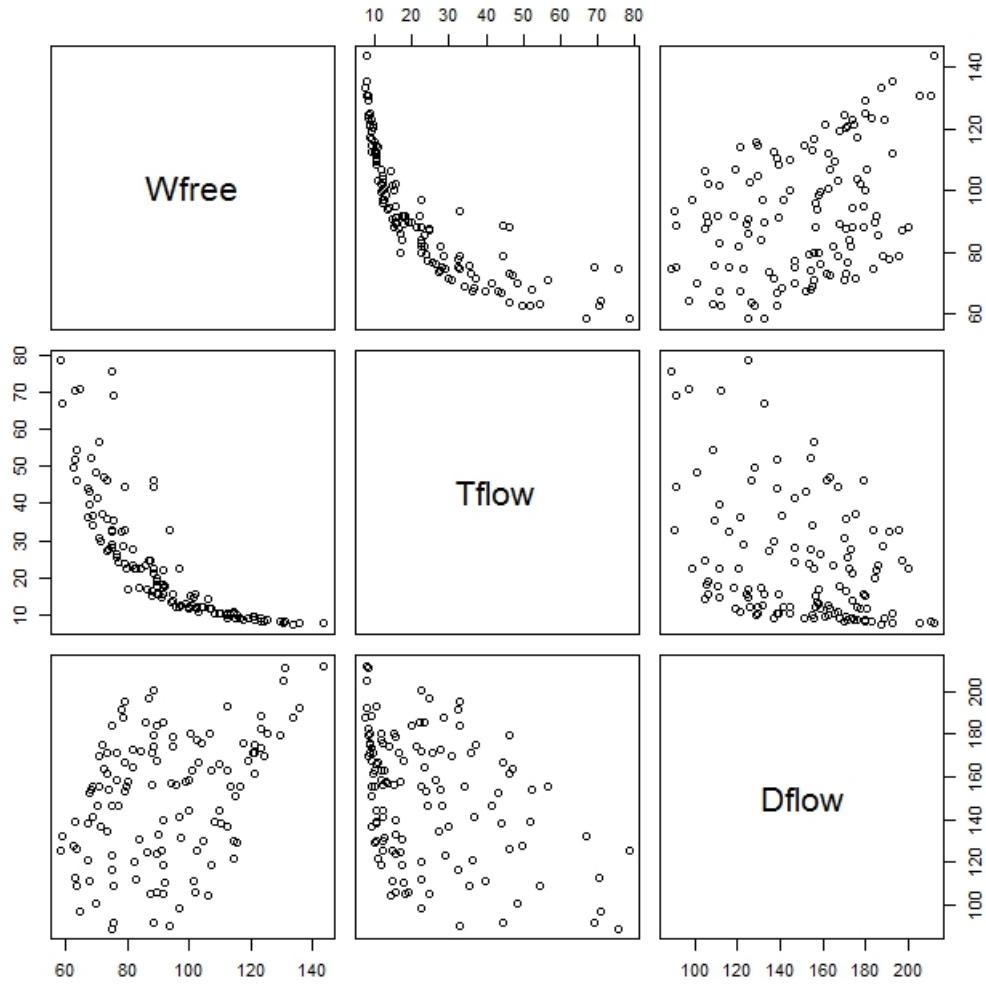


Figura 5.1: Gráfico de dispersão dos dados

de Spearman, ($r_s = -0.955$), entre essas as variáveis revelam uma correlação negativa significativa e forte. Ou seja, tipicamente, à medida que *Wfree* aumenta *Tflow* diminui.

5.2 Uma primeira abordagem de modelação

Numa primeira abordagem, admitindo que devemos sempre começar considerando o modelo mais simples para os dados que dispomos, considere-se então o seguinte modelo linear para esta relação:

$$Tflow = \beta_0 + \beta_1 Wfree \quad (5.1)$$

onde $\beta = (\beta_0, \beta_1)$ é um vector. Para obter estimativas dos coeficientes β_0 e β_1 , aplica-se a teoria clássica dos modelos de regressão linear gaussianos, obtendo-se os resultados apresentados na Tabela 5.5. Obtem-se um valor estimado para β , $\hat{\beta} = (73.392, -0.554)$, com desvio padrão para cada um dos coeficientes dado por $(3.139, 0.033)$. Como os testes t

para ambos os coeficientes são significativos, com valores $p < 0.01$, rejeitamos a hipótese de que $\beta_0 = 0$ e $\beta_1 = 0$. Admitindo os pressupostos do modelo linear gaussiano diríamos que o modelo (5.1) é significativo. A equação do modelo de regressão ajustado será, segundo a Tabela 5.5:

$$Tflow = 73.392 - 0.554Wfree \quad (5.2)$$

Na Figura 5.2 exibimos o diagrama de dispersão dos dados ($Wfree, Tflow$), e este modelo,

<i>Variável Dependente:</i>	
	Tflow
Wfree	-0.554*** (0.033)
Constant	73.392*** (3.139)
Observations	118
R ²	0.713
Adjusted R ²	0.710
Residual Std. Error	6.845 (df = 116)
F Statistic	287.944*** (df = 1; 116)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

Tabela 5.5: Resultados obtidos para o modelo de regressão linear (5.1)

o que nos revela falhas na adaptação do modelo (5.2) aos dados. Note-se que a $Tflow$ em função de $Wfree$ é decrescente, como foi visto pelos coeficientes de correlação, e que o histograma da Figura (5.3) revela alguns problemas com a distribuição dos resíduos quando comparados com distribuição normal. Portanto é necessário considerar alguns remédios tais como transformações e adição de variáveis para obter um modelo da média que melhor se adapte aos dados. Para tal considere-se a regra de Tukey-Mosley referida na secção 4.6, notando que a variável $Tflow$ é decrescente. Podemos aplicar a transformação logarítmica a esta variável, obtendo o seguinte modelo teórico

$$\log(Tflow) = \beta_0 + \beta_1 Wfree \quad (5.3)$$

onde $\beta = (\beta_0, \beta_1)$ é um vector. Aplicando novamente o modelo de regressão linear gaussiano aos dados, usando o modelo (5.3), obtemos os resultados descritos na Tabela (5.6), coluna (2). Temos então que o valor estimado para β é $\hat{\beta} = (5.482, -0.028)$, com desvio padrão para cada um dos coeficientes dado por (0.099, 0.001). Como os testes t para ambos os coeficientes são significativos, com valores $p < 0.01$ rejeitamos a hipótese de que $\beta_0 = 0$ e $\beta_1 = 0$, Podemos dizer que o modelo é significativo e a equação do modelo de regressão ajustado é

$$\log(Tflow) = 5.482 - 0.028Wfree. \quad (5.4)$$

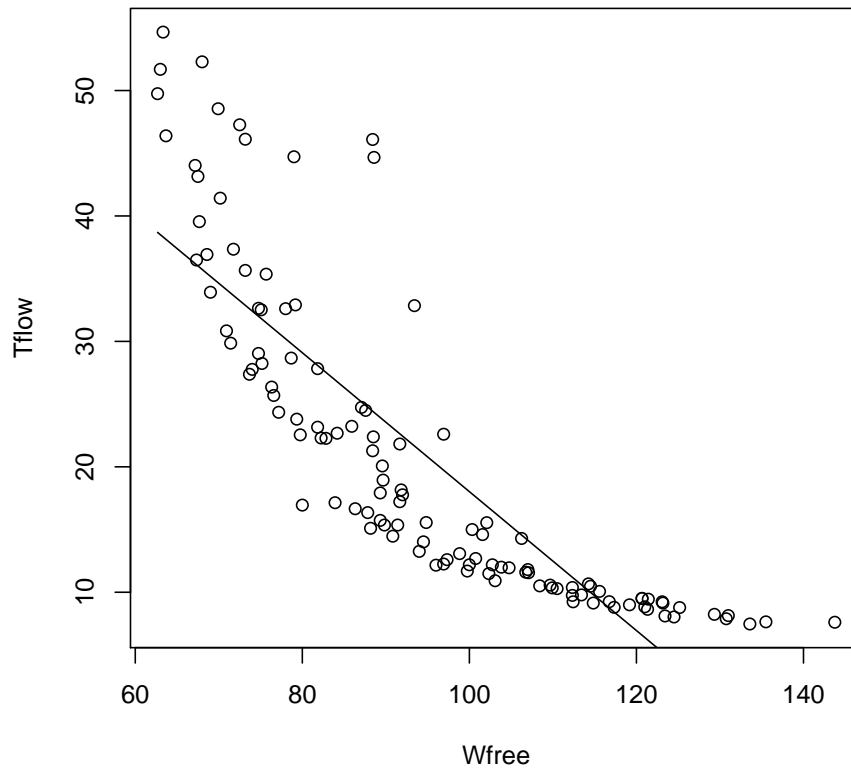


Figura 5.2: Gráfico de dispersão dos dados e a recta de regressão definida pelo modelo (5.2)

Note-se que, segundo a Tabela 5.6, pela comparação entre os modelos (5.2) e (5.4) temos uma considerável melhoria do valor do coeficiente R^2 ajustado, logo o modelo (5.4) explica melhor a variação dos dados. De notar, pela observação das Figuras 5.4 e 5.5, que a distribuição dos resíduos do modelo (5.4) aproxima-se da distribuição normal. No entanto, como se pode verificar pelas Figuras 5.3 e 5.7, o ajustamento da recta de regressão não é ainda um completo sucesso, o que nos leva a questionar que possa faltar uma variável ao modelo.

Observando a Figura 5.8 concluímos que a distribuição dos resíduos não é aleatória em torno de zero, o que é um prenúncio de heterocedasticidade. A figura apresenta um padrão curvo, representado pela linha a vermelho, no gráfico dos valores ajustados contra os resíduos. Acrescente-se que o valor da estatística teste de Tuckey para a não aditividade é 5.576 com o valor-p do teste t igual a 0 (ver Tabela 5.7). O que reforça a ideia de falta de ajustamento do modelo (5.4) aos dados. Considerando o que foi discutido sobre gráficos CERES, e reparando que a curvatura que os dados têm em relação à variável $Wfree$ é positiva, faz sentido acrescentar a variável $(Wfree)^2$ ao modelo (5.3). Isto significa que o modelo teórico passa a ser:

$$\log(Tflow) = \beta_0 + \beta_1 Wfree + \beta_2 (Wfree)^2 \quad (5.5)$$

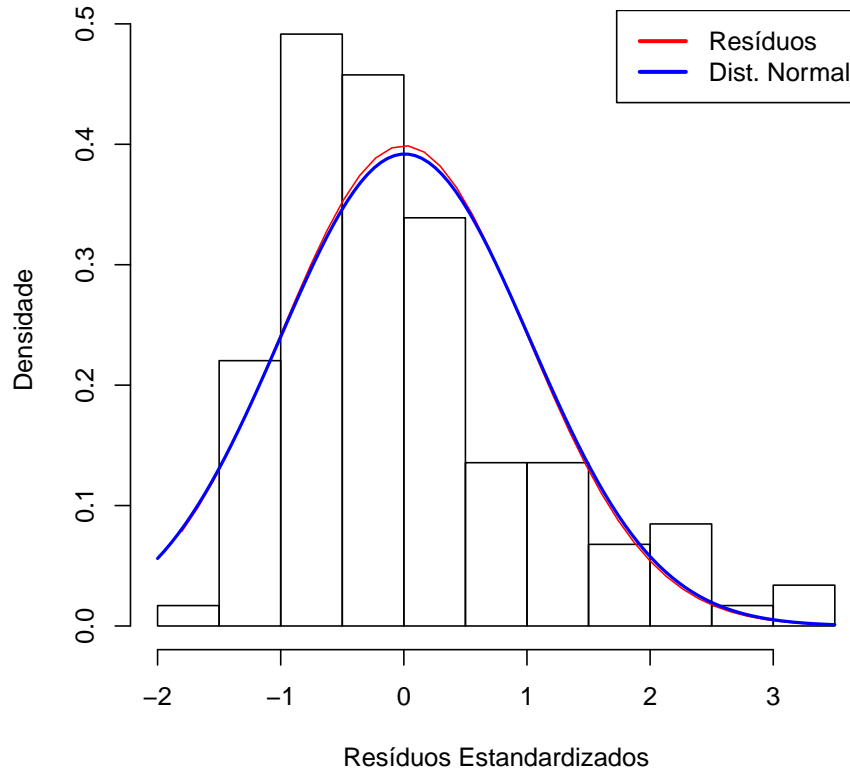


Figura 5.3: Histograma dos resíduos do modelo (5.2)

5.3 Modelo log-polinomial

Estude-se então o modelo (5.5) usando a regressão linear gaussiana, e comparando com o modelo (5.4). Obtemos a Tabela 5.8, o que implica que o modelo a considerar para os dados deste estudo é

$$\log(Tflow) = 7.809 - 0.077 Wfree + 0.0003(Wfree)^2, \quad (5.6)$$

notando-se uma melhoria ligeira do valor do coeficiente de correlação R^2 ajustado e os coeficientes de regressão sendo todos significativos, pois o seu valor- p , em todos eles, é inferior a 0.01 (ver Tabela 5.8). Observando a Figura 5.9 ficamos com a impressão de que existe uma certa diferença de variância entre valores elevados e baixos de $\log Tflow$. Ora, fazendo a representação gráfica dos resíduos estandarizados do modelo (5.6) (ver Figura 5.10), verifica-se que não existe simetria na distribuição dos valores dos resíduos em torno de zero, tanto para os valores ajustados como para as variáveis independentes do modelo (5.5).

Executando os testes mais frequentes para a homocedasticidade obtemos a Tabela 5.10, o que nos leva a rejeitar a hipótese de homocedasticidade. Portanto é necessário modelar a variância. Considerando o que foi discutido na Secção 4.8, e lembrando que:

	<i>Variável Dependente:</i>	
	Tflow	log(Tflow)
	(1)	(2)
Wfree	-0.554*** (0.033)	-0.028*** (0.001)
Constant	73.392*** (3.139)	5.482*** (0.099)
Observations	118	118
R ²	0.713	0.860
Adjusted R ²	0.710	0.859
Residual Std. Error (df = 116)	6.845	0.217
F Statistic (df = 1; 116)	287.944***	710.870***
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01		

Tabela 5.6: Resultados do modelo de regressão linear (5.3), coluna(2), e comparação com o modelo (5.1), coluna (1)

	Estatística	$Pr(> t)$
Wfree	5.576	0
Tukey	5.576	0

Tabela 5.7: Tabela do teste de curvatura para os resíduos do modelo (5.4)

- o teste de Spearman para a homocedasticidade revela uma associação entre os resíduos e a variável independente $Wfree$;
- a presença de padrões que os resíduos do modelo apresentam contra a variável $Wfree$, principalmente tanto na Figura 5.10, onde se pode observar um possível padrão em funil semelhante a uma heterocedasticidade multiplicativa, como na Figura 5.13 (onde se pode observar um padrão linear);

podemos concluir que os modelos seguintes são razoáveis para a variância:

$$\log(\epsilon^2) = \gamma_0 + \gamma_1 Wfree \quad (5.7)$$

ou

$$|\epsilon| = \gamma_0 + \gamma_1 Wfree, \quad (5.8)$$

onde ϵ é o vector dos resíduos e ϵ^2 é o vector do quadrado dos resíduos. Aplicando o método dos mínimos quadrados a ambos os modelos obtemos a Tabela 5.11. Logo, temos que os

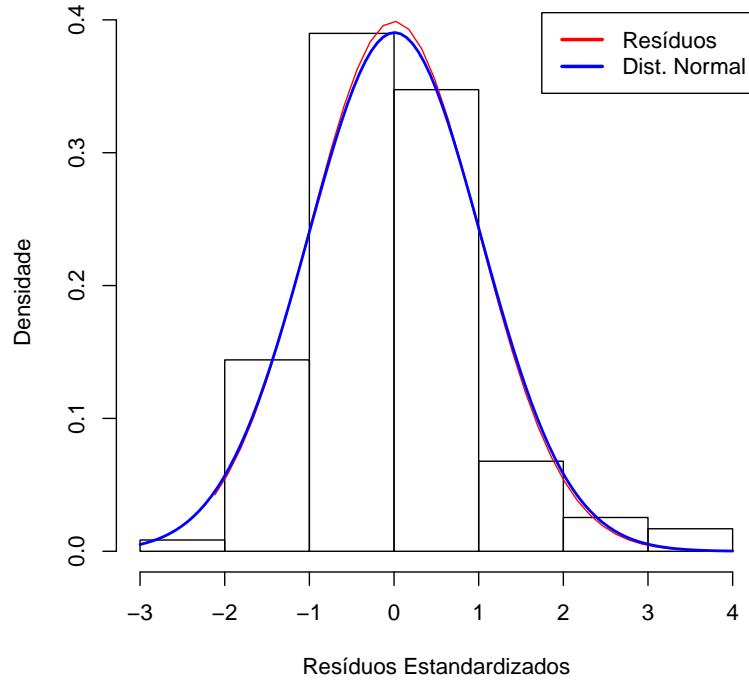


Figura 5.4: Histograma dos resíduos do modelo (5.4)

modelos de variância a considerar são:

$$\log(\epsilon^2) = -0.794 - 0.047 W_{free}, \quad (5.9)$$

$$|\epsilon| = 0.323 - 0.002 W_{free}. \quad (5.10)$$

Apesar do modelo expresso na equação (5.9) apresentar um valor R^2 ajustado mais significativo, pela tabela, não nos permite excluir o modelo (5.10), pois o que se pretende é obter uma melhor estimativa de β e não um modelo mais eficiente para a variância.

Estimação dos parâmetros e sua significância

Aplicando o método dos mínimos quadrados ponderados usando ambos os modelos de variância obtemos os resultados expostos na Tabela 5.12, que nos leva a concluir que a melhor opção para a modelação da variância é o modelo (5.10), pois apresenta um valor de R^2 ajustado mais significativo, e apresentar valores de $pseudoR^2 = 0.88902$ contra um valor de $pseudoR^2 = 0.88889$ para o modelo (5.9). O modelo a considerar é portanto (ver Tabela 5.12)

$$\log(T_{flow}) = 8.132 - 0.083 W_{free} + 0.0003 (W_{free})^2 \quad (5.11)$$

Apesar de não revelar grande melhoria no coeficiente R^2 ajustado, com este modelo temos a questão da heterocedasticidade resolvida, como se pode constatar pela Figura 5.13, onde os

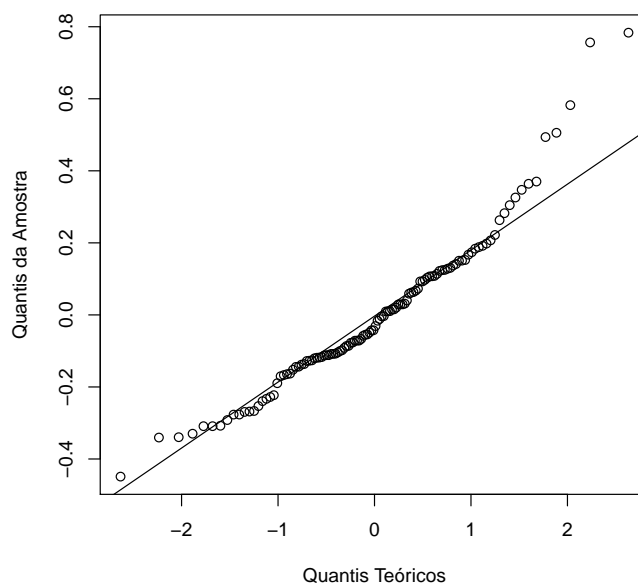


Figura 5.5: Gráfico QQ dos resíduos do modelo (5.4)

resíduos de Pearson não revelam qualquer padrão em relação a qualquer uma das variáveis independentes ou mesmo em relação aos valores ajustados. Comparando também as Figuras 5.9 e 5.14, parece-nos que se obtém uma melhor estimação para os intervalos de confiança dos valores previstos no modelo (5.11)

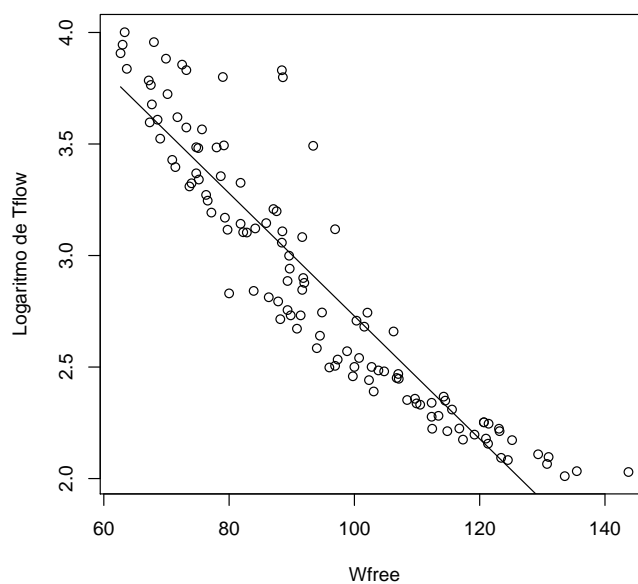


Figura 5.6: Gráfico de dispersão dos dados e recta de regressão definida pelo modelo (5.4)

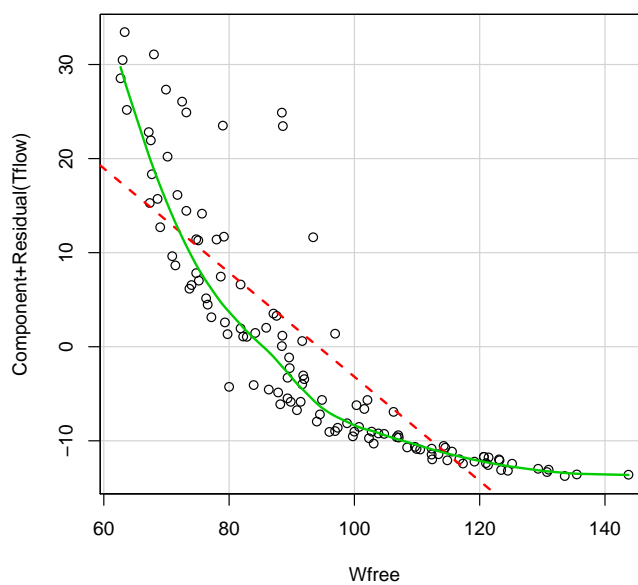


Figura 5.7: Gráfico de Componente+Resíduos do modelo (5.4)

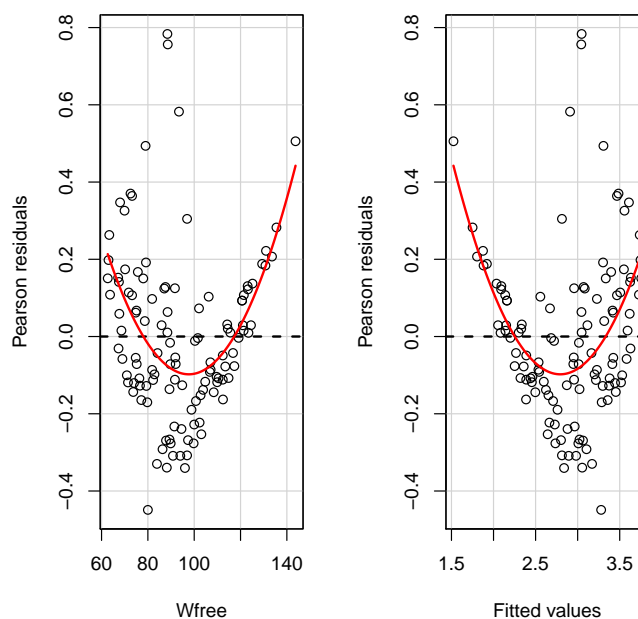


Figura 5.8: Gráfico dos resíduos de Pearson do modelo (5.4)

	<i>Variável Dependente::</i>	
	log(Tflow)	
	(1)	(2)
Wfree	-0.028*** (0.001)	-0.077*** (0.009)
I(Wfree^2)		0.0003*** (0.00005)
Constant	5.482*** (0.099)	7.809*** (0.427)
Observations	118	118
R ²	0.860	0.890
Adjusted R ²	0.859	0.888
Residual Std. Error	0.217 (df = 116)	0.193 (df = 115)
F Statistic	710.870*** (df = 1; 116)	463.187*** (df = 2; 115)

Note:

*p<0.1; **p<0.05; ***p<0.01

Tabela 5.8: Resultados do modelo de regressão linear (5.5), coluna(2) e comparação com o modelo (5.3), coluna (1)

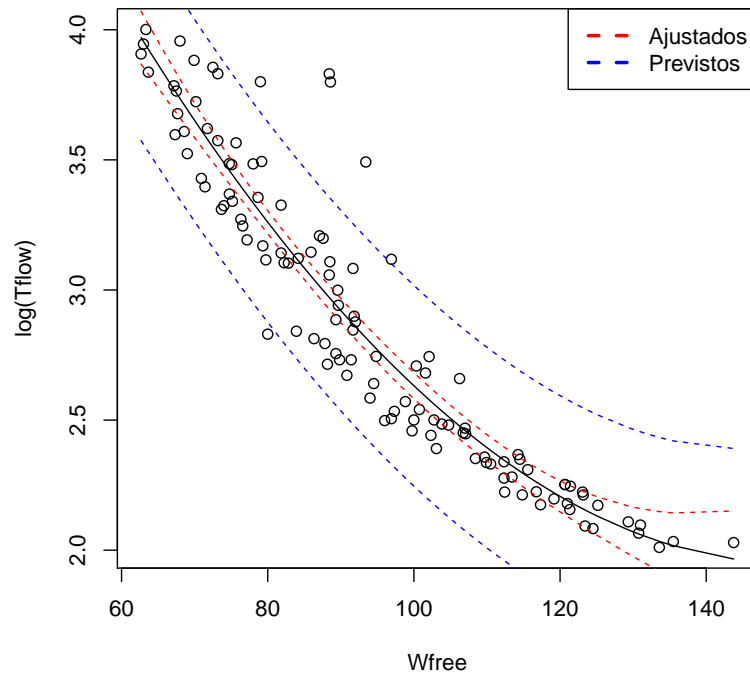


Figura 5.9: Gráfico de dispersão dos dados com a curva de regressão definida pelo modelo (5.6), com os respectivos intervalos de confiança para os valores ajustados da regressão e para os valores previstos.

	Estatística	$Pr(> t)$
W_{free}	0.425	0.672
$I(W_{free}^2)$	0.865	0.389
Tukey test	-0.893	0.372

Tabela 5.9: Tabela do teste de curvatura para os resíduos do modelo (5.6)

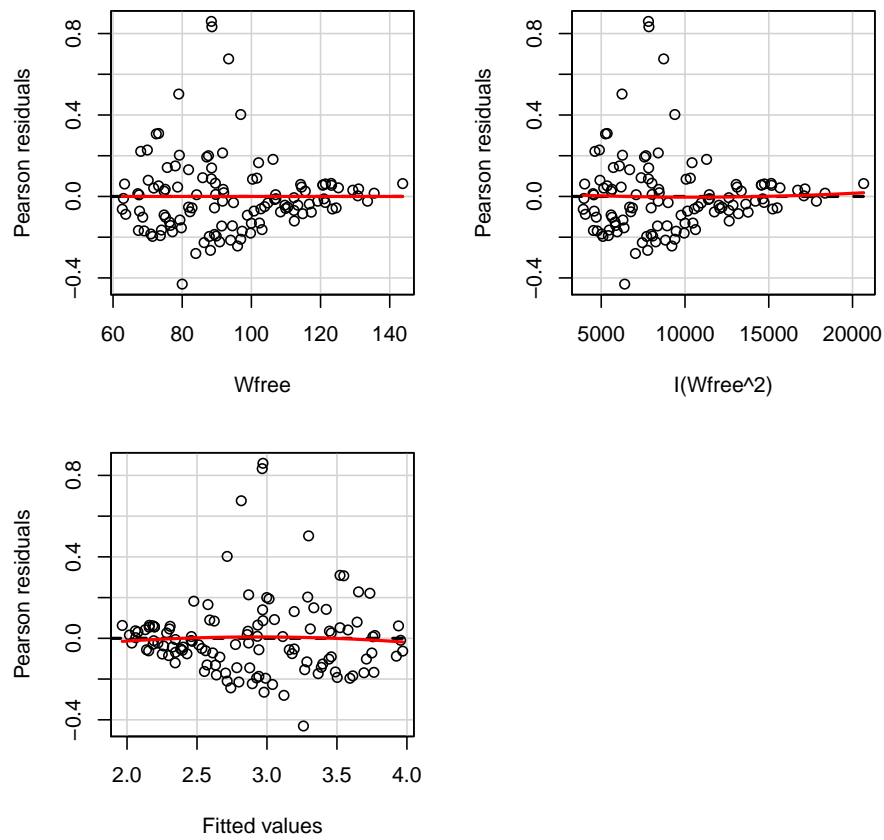


Figura 5.10: Gráfico dos resíduos de Pearson do modelo (5.6)

Nome do Teste	Fórmula da Variância	Estatística	Valor-p	H_0
Spearman	\widehat{Wfree}	-4.2352	0	Homocedástico
Score	$\log(Tflow)$	4.285015	0.38444974	Homocedástico
Score	\widehat{Wfree}	7.813147	0.005186749	Homocedástico
Breusch-Pagan		5.2259	0.07332	Homocedástico
Levene		4.073798	4.247016×10^{-05}	Homocedástico
Koenker		21.504	0	Homocedástico
White		105.06	0	Homocedástico
Goldfeld-Quandt		0.35737	0	Homocedástico
Carapeto & Holt		0.95157	0.5005	Heterocedástico
Glejser	\widehat{Wfree}^{-1}	2.1924	0.03035	Homocedástico
Glejser	$\widehat{Wfree}^{-0.5}$	2.3912	0.0184	Homocedástico
Glejser	$\widehat{Wfree}^{0.5}$	-2.7594	0.006732	Homocedástico
Glejser	\widehat{Wfree}^1	-2.9222	0.004179	Homocedástico
Harvey		781.153	< 0.001	Homocedástico
Park	$\exp(\widehat{Wfree})$	-4.654	8.71×10^{-6}	Homocedástico

Tabela 5.10: Testes de homocedasticidade para o modelo (5.6)

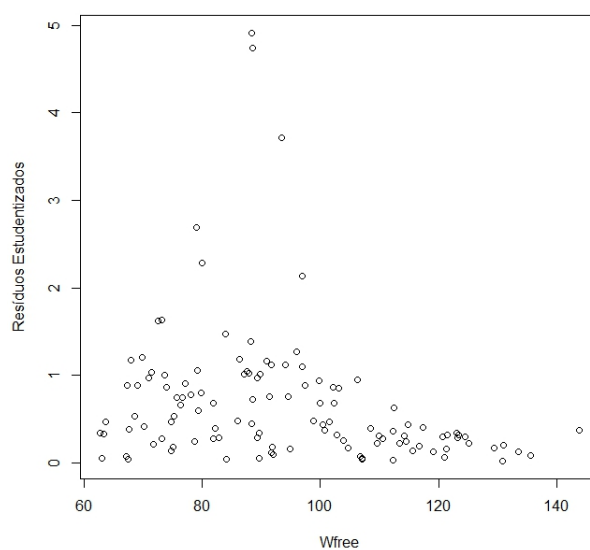


Figura 5.11: Gráfico dos resíduos estudentizados para aferir da forma da heterocedasticidade, onde os eixos são respectivamente na horizontal a variável independente W_{free} e na vertical o valor absoluto dos resíduos estudentizados

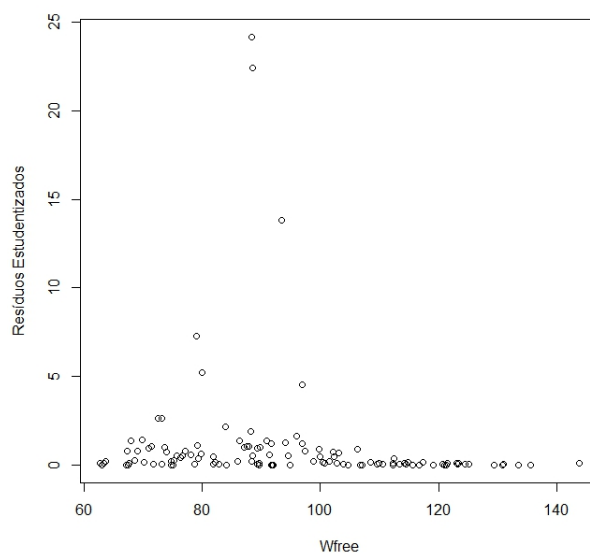


Figura 5.12: Gráfico dos resíduos estudentizados para aferir da forma da heterocedasticidade, onde os eixos são respectivamente na horizontal a variável independente W_{free} e na vertical o quadrados dos resíduos estudentizados

	<i>Variável Dependente:</i>	
	$\log(res^2)$	$ res $
	(1)	(2)
Wfree	-0.047*** (0.010)	-0.002*** (0.010)
Constant	-0.794 (0.965)	0.323 (0.965)
Observations	118	118
R ²	0.157	0.079
Adjusted R ²	0.150	0.071
Residual Std. Error (df = 116)	2.104	0.138
F Statistic (df = 1; 116)	21.662***	9.988***
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01		

Tabela 5.11: Resultados do modelo de regressão linear (5.7), coluna (1), e (5.8), coluna (2)

	<i>Variável Dependente:</i>		
	$\log(Tflow)$		
	(1)	(2)	(3)
Wfree	-0.077*** (0.009)	-0.084*** (0.007)	-0.084*** (0.006)
I(Wfree ²)	0.0003*** (0.00005)	0.0003*** (0.00003)	0.0003*** (0.00003)
Constant	7.809*** (0.427)	8.157*** (0.363)	8.132*** (0.305)
Observations	118	118	118
R ²	0.890	0.907	0.926
Adjusted R ²	0.888	0.905	0.924
Residual Std. Error (df = 115)	0.193	2.012	1.414
F Statistic (df = 2; 115)	463.187***	558.067***	717.308***
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01			

Tabela 5.12: Resultados do modelo de regressão linear (5.5) usando o modelo de variância (5.9), coluna(2), e comparação com o modelo (5.5), coluna (1), e (5.5) usando o modelo de variância (5.10), coluna(3)

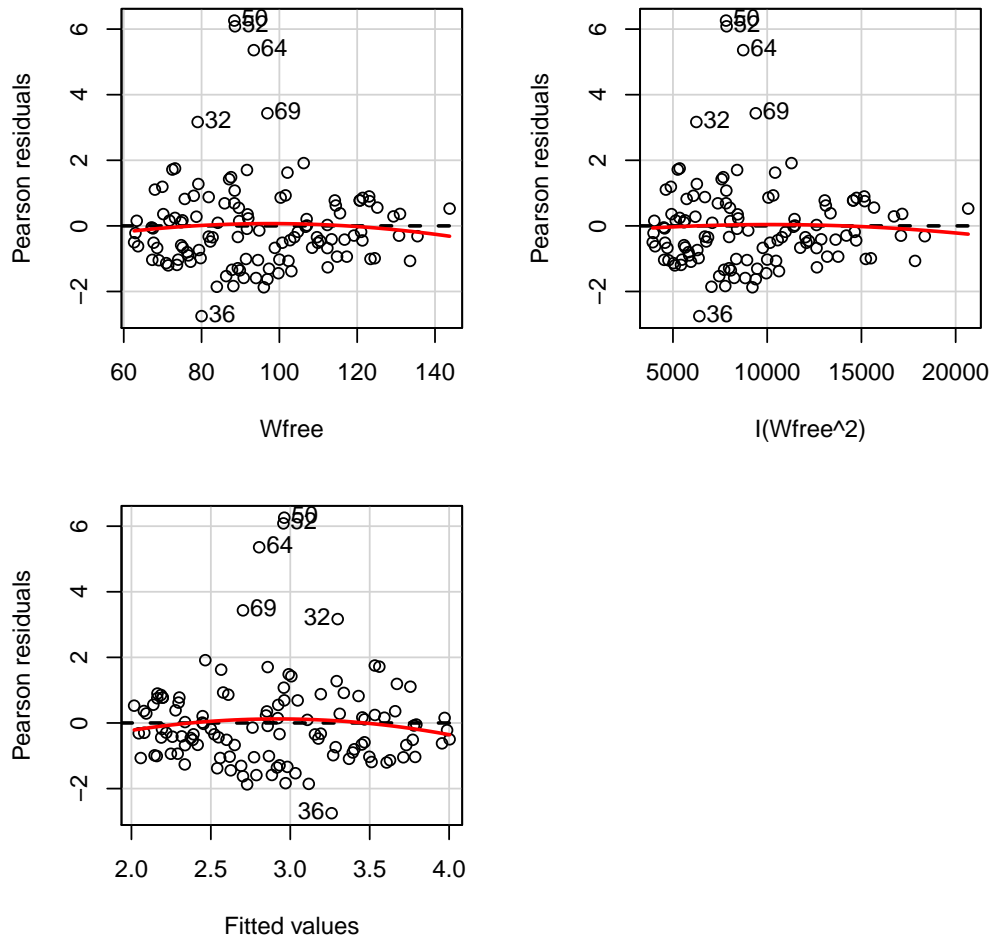


Figura 5.13: Gráfico das variáveis do modelo (5.11) contra os resíduos de Pearson.

	Estatística	$Pr(> t)$
Wfree	-2.128	0.035
I(Wfree^2)	-2.138	0.035
Tukey test	-2.121	0.034

Tabela 5.13: Tabela do teste de curvatura para os resíduos do modelo (5.11)

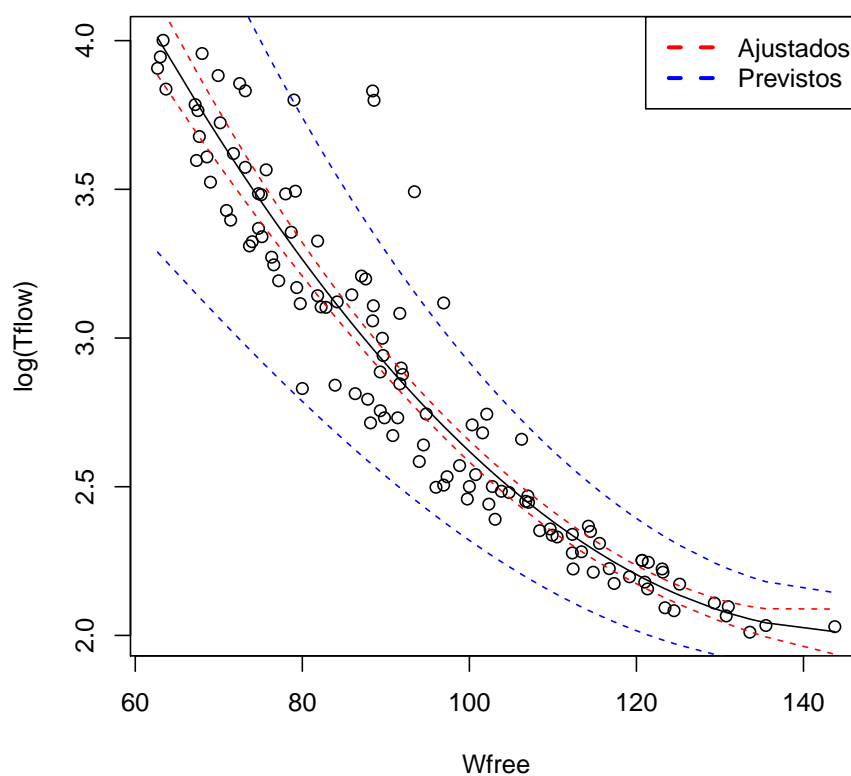


Figura 5.14: Gráfico de dispersão dos dados com a curva de regressão definida pelo modelo (5.11), com os respectivos intervalos de confiança para os valores ajustados da regressão e para os valores previstos.

Verificação dos pressupostos do modelo

O modelo definido em (5.11) só será adequado se validados todos os pressupostos. Vamos nesta subsecção fazer uma análise desses pressupostos.

Normalidade dos resíduos

A partir da análise das Figuras 5.15 e 5.16, podemos concluir que as observações se aproximam da recta sem nenhum afastamento sistemático, embora haja uma pequena percentagem de pontos com mau comportamento, pelo que somos levados a concluir que os resíduos são normalmente distribuídos. Alguns valores se afastam da recta poderão corresponder a outliers. Apesar de (ver Tabela 5.14) os testes de normalidade dos resíduos rejeitarem

Teste	Estatística	Valor-p	H_0
Kolmogorov-Smirnov	0.38308	1.776×10^{-15}	Normais
Shapiro-Wilk	0.84043	5.925×10^{-10}	Normais

Tabela 5.14: Testes de normalidade dos resíduos

a hipótese de normalidade, está questão da normalidade dos resíduos não se coloca pois estes são os resíduos observados e não os reais. São uma amostra de uma população que assumimos normal, e o Teorema do Limite Central e a Lei dos Grandes Números garantem a normalidade destas observações.

Autocorrelação dos resíduos

Teste	Autocorrelação	Estatística	Valor-p	H_0
Durbin Watson	0.05190805	1.893632	0.434	$\rho = 0$

Tabela 5.15: Teste de autocorrelação dos resíduos

Aplicando o teste de Durbin-Watson aos resíduos (ver Tabela 5.15) somos levados a concluir que os resíduos são independentes.

Outliers

Aplicando testes à existência de outliers obtém-se os resultados expressos na Tabela 5.16 e representados na Figura 5.17. Nesta Figura pode identificar-se os pontos que são outliers. As observações 50, 52 e 64, numa primeira aplicação do teste revelam um valor-p de Boferroni estatisticamente significativo, logo deve-se excluir esses pontos na construção do modelo de regressão. Uma segunda aplicação do mesmo teste revela uma última observação, 69, com um valor p significativo. Retirando estes pontos obtemos os resultados que constam da

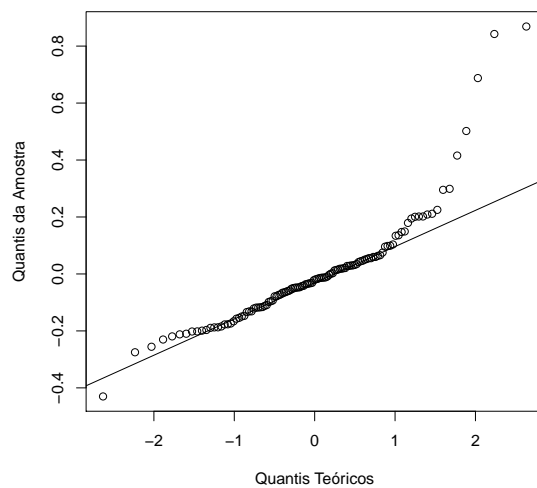


Figura 5.15: Gráfico QQ normais para os resíduos do modelo (5.11)

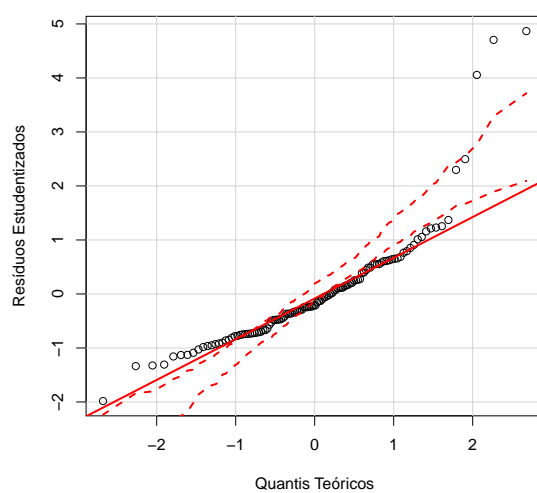


Figura 5.16: Gráfico dos Resíduos Estudatizados com os intervalos de confiança para o modelo (5.11)

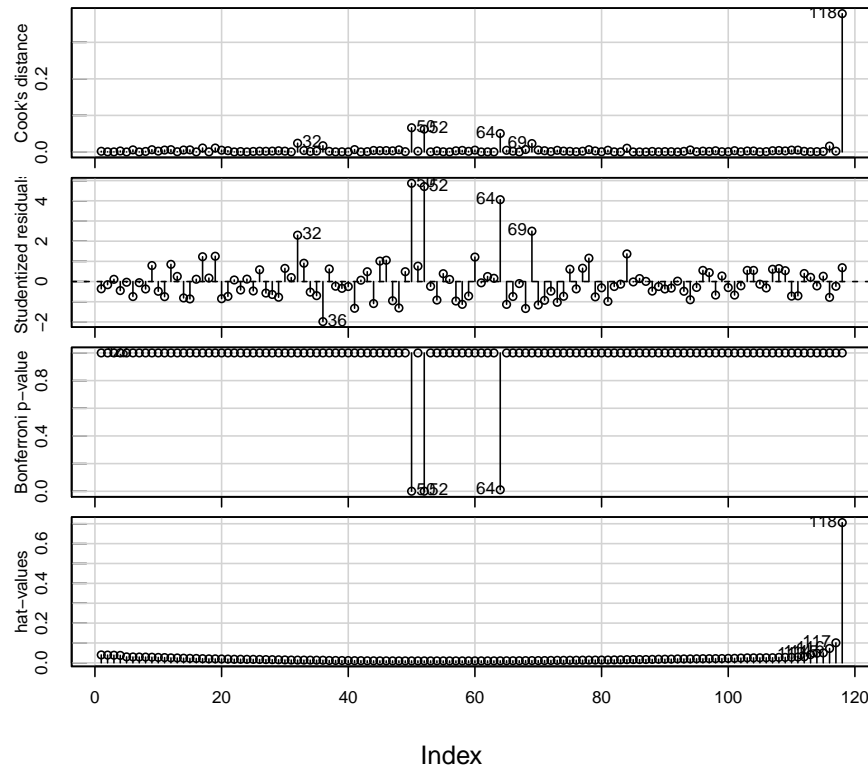


Figura 5.17: Gráficos de diagnóstico para as estatísticas do modelo (5.11)

Tabela 5.18 o que nos revela uma melhoria do valor de R^2 ajustado significativa, levando a optar por retirar as observações citadas (ver Tabela 5.17). Obtem-se o modelo:

$$\log(Tflow) = 8.136 - 0.085 Wfree + 0.0003(Wfree)^2. \quad (5.12)$$

Observando as Figuras 5.18 e 5.19 e comparando com as Figuras 5.15 e 5.16 conclui-se que a exclusão dos outliers melhora consideravelmente a normalidade dos resíduos.

Pontos de alavanca

Na Figura 5.20 representamos as distâncias de Cook e outras medidas de diagnóstico para o modelo (5.12). Pela observação desta Figura conclui-se que existe uma observação, a 114, com um valor de alavancagem que apesar de estar dentro do que na prática se considera um valor razoável, o ponto é excluído, pois, no contexto do problema, esta observação é para um valor de $Dflow=212.00$ mm, um valor considerado elevado, pois esta amassadura fica no intervalo para os quais se começa a verificar exsudação. Analisando a Tabela 5.19, constatamos que apesar de perdermos um pouco de qualidade de ajuste, pois o valor de

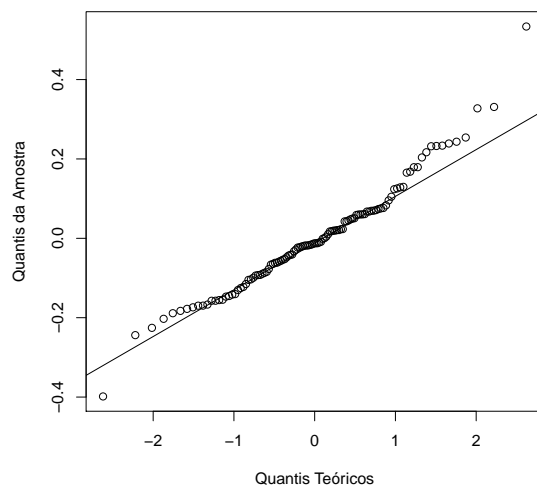


Figura 5.18: Gráfico QQ normais para os resíduos do modelo (5.12)

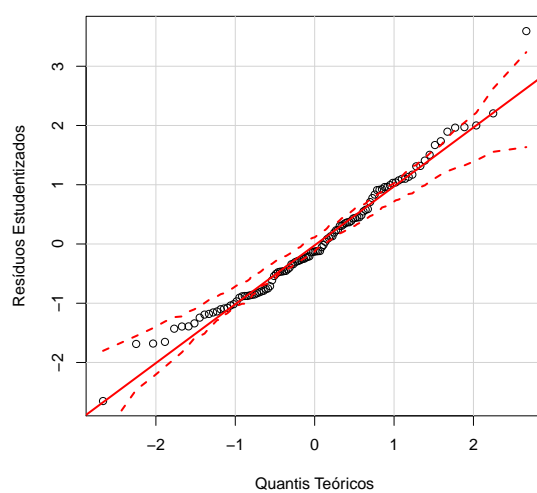


Figura 5.19: Gráfico dos Resíduos Estudentizados com os intervalos de confiança para o modelo (5.12)

Nº do Resíduo	Valor do Resíduo	Valor-p não ajustado	Bonferroni-p
50	4.867163	3.6687e-06	0.00043290
52	4.704741	7.1916e-06	0.00084861
64	4.055875	9.1724e-05	0.01082300
69	3.673557	0.00036979	0.042526

Tabela 5.16: Testes aos Outliers

Nº da Observação	Water	Wfree	Tflow	Dflow
50	0.32	88.42	46.1	179.25
52	0.36	88.58	44.67	91.5
64	0.37	93.42	32.85	90.25
69	0.38	96.92	22.6	98.5

Tabela 5.17: Valores das variáveis excluídas na construção do modelo (5.12)

	<i>Variável Dependente:</i>	
	log(Tflow)	
	(1)	(2)
Wfree	-0.084*** (0.006)	-0.085*** (0.004)
I(Wfree ²)	0.0003*** (0.00003)	0.0003*** (0.00002)
Constant	8.132*** (0.305)	8.146*** (0.214)
Observations	118	114
R ²	0.926	0.961
Adjusted R ²	0.924	0.960
Residual Std. Error	1.414 (df = 115)	0.990 (df = 111)
F Statistic	717.308*** (df = 2; 115)	1,372.339*** (df = 2; 111)

Note:

*p<0.1; **p<0.05; ***p<0.01

Tabela 5.18: Resultados do modelo de regressão linear (5.12) retirados os outliers, coluna(2), e comparação com o modelo (5.11), coluna (1)

R^2 ajustado diminui, retirando a observação 114, obtemos um modelo que reflecte melhor o pretendido para o tipo de pasta de betão em estudo. Ficamos, então, com o seguinte

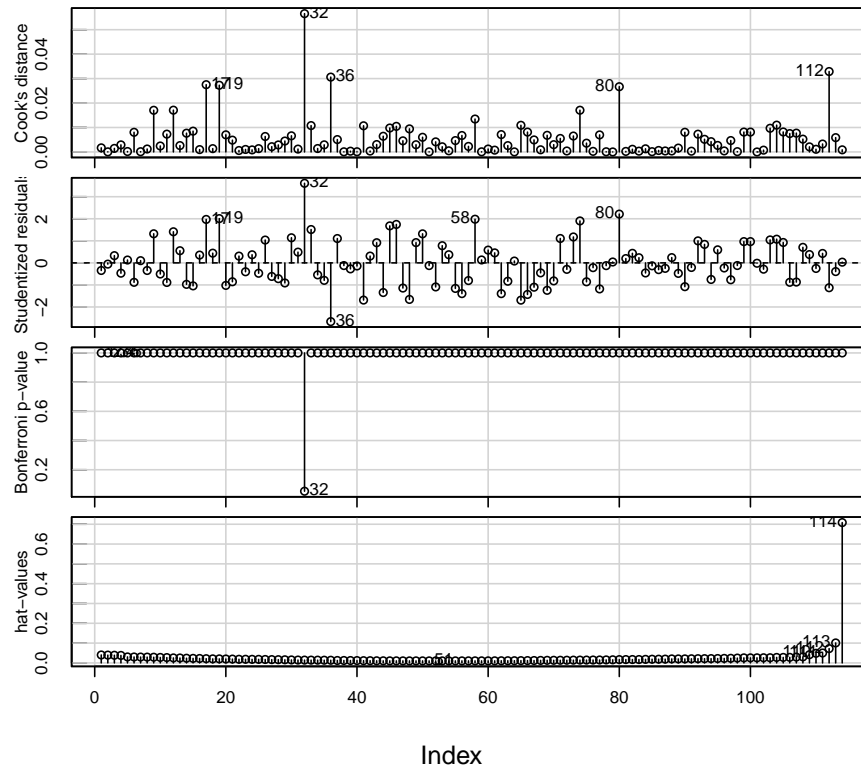


Figura 5.20: Gráficos de diagnóstico para as estatísticas do modelo (5.12)

modelo:

$$\log(Tflow) = 8.151 - 0.085 Wfree + 0.0003(Wfree)^2 \quad (5.13)$$

	<i>Dependent variable:</i>	
	log(Tflow)	
	(1)	(2)
Wfree	-0.085*** (0.004)	-0.085*** (0.005)
I(Wfree^2)	0.0003*** (0.00002)	0.0003*** (0.00003)
Constant	8.146*** (0.214)	8.140*** (0.276)
Observations	114	113
R ²	0.961	0.956
Adjusted R ²	0.960	0.955
Residual Std. Error	0.990 (df = 111)	0.994 (df = 110)
F Statistic	1,372.339*** (df = 2; 111)	1,199.937*** (df = 2; 110)
<i>Note:</i>		*p<0.1; **p<0.05; ***p<0.01

Tabela 5.19: Resultados do modelo de regressão linear (5.13) uma vez retirados os pontos de influência, coluna(2), e comparação com o modelo (5.12), coluna (1)

Homocedasticidade dos resíduos

Na Figura 5.22 representamos os gráficos dos resíduos do modelo (5.13). A partir da análise gráfica desta figura parece lícito admitir que os resíduos do modelo (5.13) são homoscedásticos, uma vez que estes se distribuem de forma aleatória em torno zero e sem padrão aparente. No entanto, pela análise da Tabela 5.20, poderíamos concluir que a forma funcional do modelo estaria incorrecta, pois os testes estatísticos de curvatura têm valor- p inferior a 5%. No entanto, aqui estaríamos a cometer um erro de Tipo I pois, e atendendo à Figura 5.22 não parece existir curvatura.

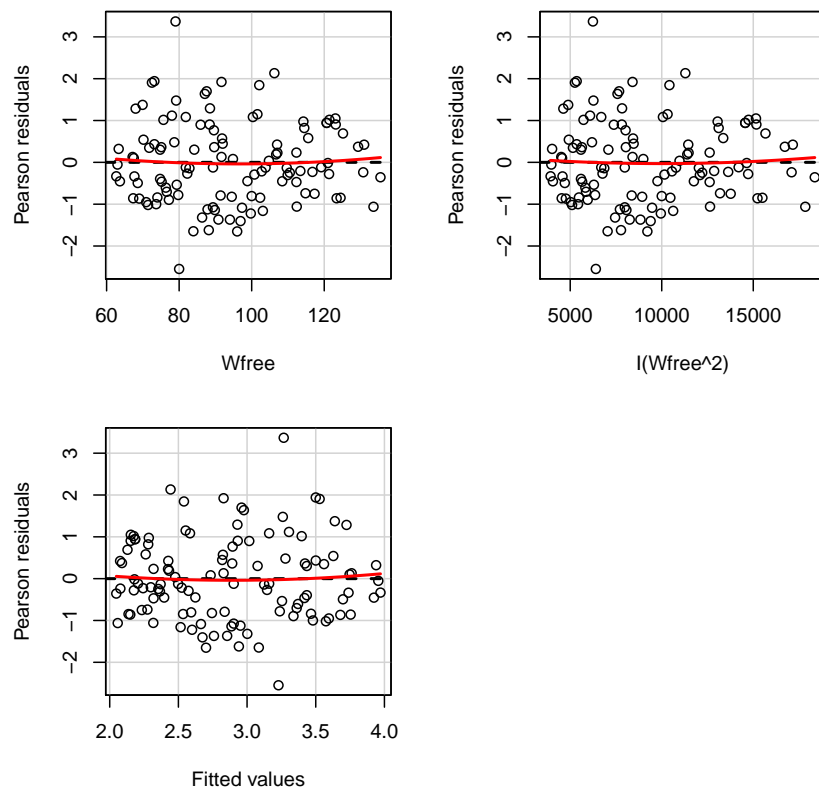


Figura 5.21: Gráfico dos resíduos de Pearson do modelo (5.13)

	Estatística	$Pr(> t)$
Wfree	2.579	0.011
I(Wfree^2)	2.675	0.009
Tukey test	2.467	0.014

Tabela 5.20: Tabela do teste de curvatura para os resíduos do modelo (5.12)

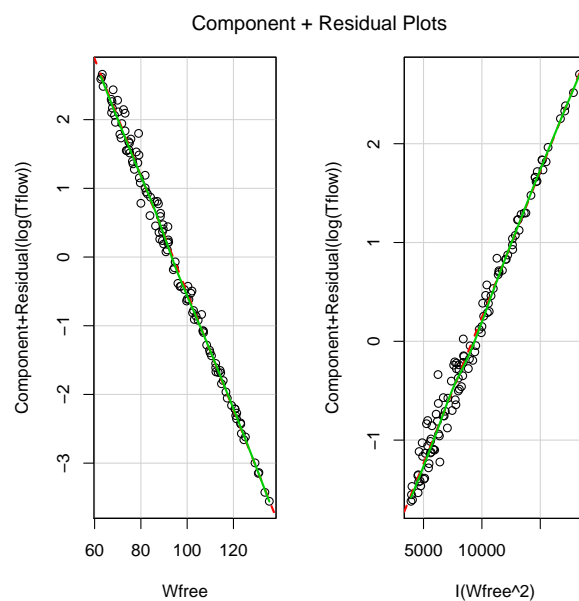


Figura 5.22: Gráfico Componente+Resíduo do modelo (5.13)

O Modelo final

Uma vez que todos os pressupostos da regressão linear simples foram validados, podemos concluir que o modelo (5.13) é adequado, representando devidamente os dados. Este modelo está representado na Figura 5.23.

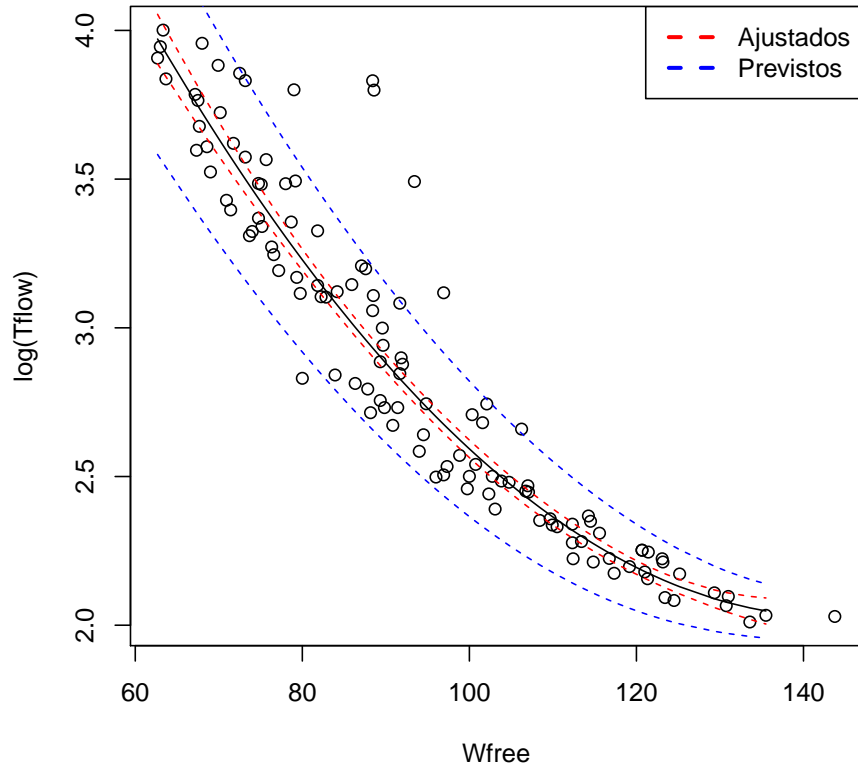


Figura 5.23: Gráfico de dispersão dos dados e curva de regressão definida pelo modelo (5.13), com os respectivos intervalos de confiança para os valores ajustados da regressão e para os valores previstos.

5.4 Modelo hiperbólico

Considere-se novamente o modelo linear

$$T_{flow} = \beta_0 + \beta_1 W_{free}, \quad (5.14)$$

onde $\beta = (\beta_0, \beta_1)$ é um vector de parâmetros. Aplicando agora a transformação de Box-Cox (ver 4.6) obtém-se a curva da Figura 5.24, que nos sugere que uma possível escolha para λ seria $\lambda = -1$, ou seja, devemos optar pela transformação inversa da resposta:

$$\frac{1}{T_{flow}} = \beta_0 + \beta_1 W_{free} \quad (5.15)$$

Aplicamos os mesmos procedimentos da secção anterior, (secção 5.3), no que diz respeito à modelação da heterocedasticidade, onde a heterocedasticidade segue um modelo multiplica-

tivo, com a variância modelada pela seguinte relação teórica:

$$\log(\sigma^2) = \gamma_0 + \gamma_1 W_{free}. \quad (5.16)$$

Encontrando os mesmos pontos de outliers e pontos de alavancagem, e obtemos, por fim, o seguinte modelo:

$$\begin{aligned} \frac{1}{T_{flow}} &= -0.116945 + 0.002189 W_{free} - 0.000003 W_{free}^2, \\ \log(\sigma^2) &= -0.7936001000 - 0.0466841100 W_{free}, \end{aligned} \quad (5.17)$$

através de uma regressão pelo método dos mínimos quadrados ponderados. Os resultados obtidos são explanados na Tabela (5.21), onde se compara os resultados entre este modelo, (modelo 5.17) coluna (2), e o modelo (5.13), coluna(1). Da análise desta tabela conclui-se que o ajustamento do modelo (5.17) aos dados é melhor do que o modelo (5.13), mas tomando em atenção as Figuras 5.25, 5.26 e 5.27 em comparação com as Figuras 5.18 e 5.19 conclui-se que apesar de um melhor ajustamento do modelo (5.17), o modelo (5.13) apresenta melhores resultados para a distribuição dos resíduos. O modelo (5.17) está representado na Figura 5.28 com os respectivos intervalos de confiança para os valores ajustados (linha tracejada a vermelho) e os intervalos de previsão (linha tracejada a azul).

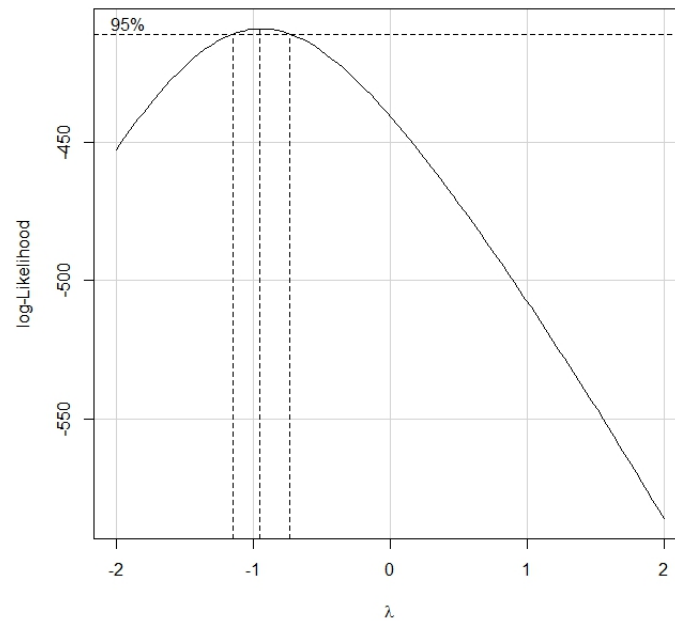


Figura 5.24: Gráfico da transformação de Box-Cox para a variável resposta do modelo (5.15).

	<i>Variável Dependente:</i>	
	log(Tflow)	1/Tflow
	(1)	(2)
Wfree	-0.085*** (0.006)	0.002189*** (0.000380)
I(Wfree^2)	-0.0003*** (0.00003)	-0.000003 (0.000002)
Constant	8.151*** (0.298)	-0.116945*** (0.019595)
Observations	113	113
R ²	0.949	0.954722
Adjusted R ²	0.948	0.953899
Residual Std. Error	1.390 (df = 110)	0.091350 (df = 110)
F Statistic	1,014.383*** (df = 2; 111)	1,159.722000*** (df = 2; 110)
<i>Note:</i>		*p<0.1; **p<0.05; ***p<0.01

Tabela 5.21: Resultados do modelo de regressão linear (5.13), coluna(1), e comparação com o modelo (5.17), coluna (2)

	<i>Variável Dependente:</i>
	$\log(res^2)$
Wfree	-0.0466841100*** (0.0100305600)
Constant	-0.7936001000 (0.9648674000)
Observations	118
R ²	0.1573532000
Adjusted R ²	0.1500891000
Residual Std. Error	2.1041280000 (df = 116)
F Statistic	21.6614800000*** (df = 1; 116)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Tabela 5.22: Resultados do modelo de regressão linear (5.16)

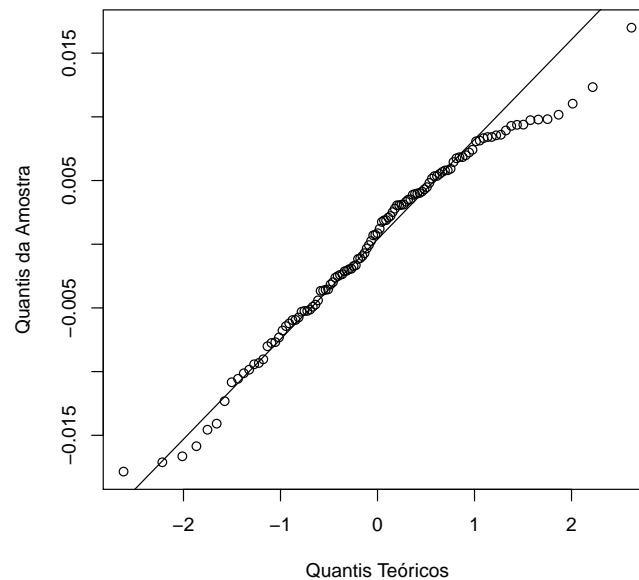


Figura 5.25: Gráfico QQ normais para o modelo (5.17)

5.5 Discussão dos resultados

Face ao exposto nas secções anteriores, a escolha do modelo mais adequado para representar a variação da variável $Tflow$ como função da variável $Wfree$ recai sobre o modelo log-polinomial, por três razões principais:

1. apresenta melhores resultados ao nível de resíduos do que outros modelos;
2. no campo em estudo é frequente o seu uso, (ver [74, 110]);
3. é de mais fácil interpretação.

Apesar de o coeficiente de determinação ser menor (ver Tabela 5.24), a diferença encontrada no ajustamento entre o modelo log-polinomial e o modelo hiperbólico não justifica um aumento na dificuldade em interpretar os resultados.

Resta-nos investigar a utilização de métodos não paramétricos, e comparar no modelo os resultados que se obtêm nomeadamente na avaliação da capacidade preditiva do modelo adoptado. Seguindo o procedimento delineado por Kuk (ver [95]) para modelar a variância, usando métodos não paramétricos, e por Rilstone usando semi-paramétricos (ver [119]), obteve-se os resultados sumarizados na Tabela 5.23, onde se compara as estimativas pelo método dos mínimos quadrados ponderados com pesos estimados pelo método dos mínimos quadrados para o modelo (5.13) (coluna (1)) e usando diferentes tipos de ponderação com estimações do modelo da média pelo método dos mínimos quadrados ponderados com pesos estimados por uma regressão quantil (coluna (2)), regressão Loess (coluna (3)), regressão por núcleos (coluna (4)), regressão Lasso (coluna (5)). Para cada um destes métodos representamos os dados com as respectivas curvas que representam o modelo, intervalos

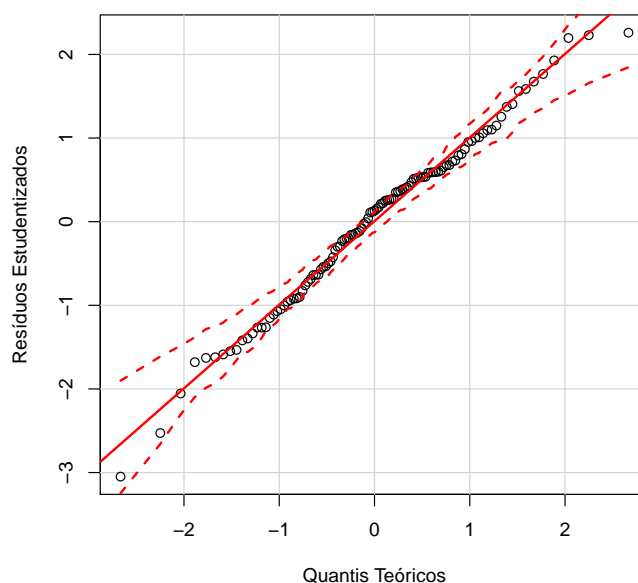


Figura 5.26: Gráfico dos resíduos estudentizados com os intervalos de confiança modelo (5.17)

de confiança para os valores ajustados e intervalos de previsão, (ver Figuras 5.29 e 5.29). A análise desses resultados permite concluir que o modelo (5.13) é o suficientemente preciso para ser usado em previsão.

A análise da Tabela 5.24, obtida usando estimadores de White para o modelo (5.6), permite concluir que não obteríamos melhores resultados usando este estimador pois os valores de R^2 ajustados são menores para estas estimativas do que para o modelo (5.13).

Variável Dependente:						
	(1)	(2)	log(Tflow)	(3)	(4)	(5)
Wfree	-0.085*** (0.005)	-0.085*** (0.005)		-0.084*** (0.006)	-0.082*** (0.004)	-0.085*** (0.005)
I(Wfree^2)	0.0003*** (0.00003)	0.0003*** (0.00003)		0.0003*** (0.00003)	0.0003*** (0.00002)	0.0003*** (0.00003)
Constant	8.140*** (0.276)	8.140*** (0.276)		8.076*** (0.290)	8.007*** (0.216)	8.140*** (0.276)
Observations	113	113		114	114	113
R ²	0.956	0.956		0.951	0.968	0.956
Adjusted R ²	0.955	0.955		0.951	0.968	0.955
Residual Std. Error	0.994 (df = 110)	0.994 (df = 110)		1.048 (df = 111)	1.042 (df = 111)	0.994 (df = 110)
F Statistic	1,199.937*** (df = 2; 110)	1,199.937*** (df = 2; 110)		1,086.687*** (df = 2; 111)	1,686.880*** (df = 2; 111)	1,199.937*** (df = 2; 110)
Note:	* p<0.1; ** p<0.05; *** p<0.01					

Tabela 5.23: Resultados do modelo de regressão linear (5.13), coluna(1), em comparação com a variância modelada por: regressão Quantil (2), regressão Loess (3), regressão por núcleos (4), regressão Lasso (5).

Variável Dependente:					
	(1)	(2)	(3)	(4)	(5)
log(Tflow)					
Wfree	-0.08595*** (0.005793)	-0.08595*** (0.005872)	-0.08595*** (0.005875)	-0.08595*** (0.005959)	-0.08595*** (0.00525)
I(Wfree^2)	0.0003014*** (0.00002902)	0.0003014*** (0.00002941)	0.0003014*** (0.00002942)	0.0003014*** (0.0000284)	0.0003014*** (0.00002969)
Constant	8.172*** (0.2772)	8.172*** (0.2809)	8.172*** (0.2812)	8.172*** (0.2854)	8.172*** (0.2836)
Observations	113	113	113	113	113
R ²	0.9423	0.9423	0.9423	0.9423	0.9423
Adjusted R ²	0.9412	0.9412	0.9412	0.9412	0.9412
Residual Std. Error	0.1357 (df = 110)	0.1357 (df = 110)	0.1357 (df = 111)	0.1357 (df = 110)	0.1357 (df = 110)
F Statistic	897.9*** (df = 2; 110)	897.9*** (df = 2; 110)	897.9*** (df = 2; 110)	897.9*** (df = 2; 110)	897.9*** (df = 2; 110)
Note: * p<0.1; ** p<0.05; *** p<0.01					

Tabela 5.24: Resultados do modelo de regressão linear usando estimadores HCCM: HC0(1), HC1 (2), HC2 (3), HC3 (4), HC4 (5).

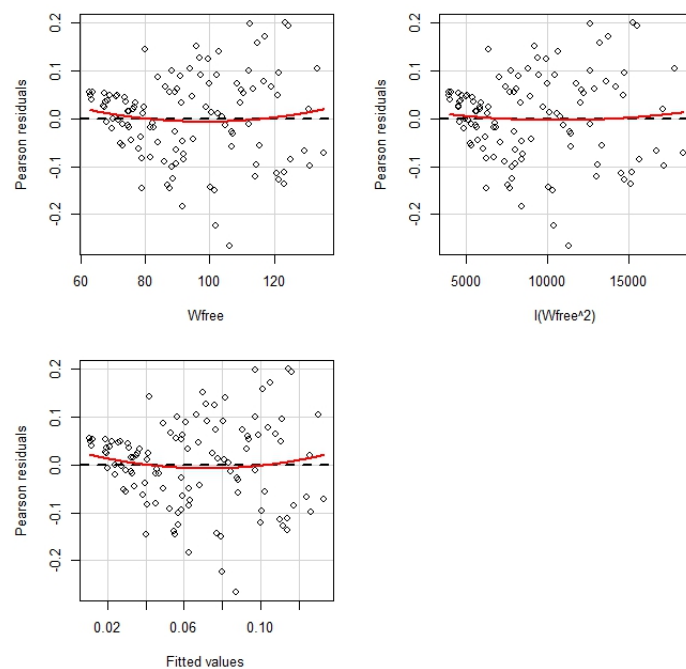


Figura 5.27: Gráfico das variáveis do modelo (5.17) contra os resíduos de Pearson.

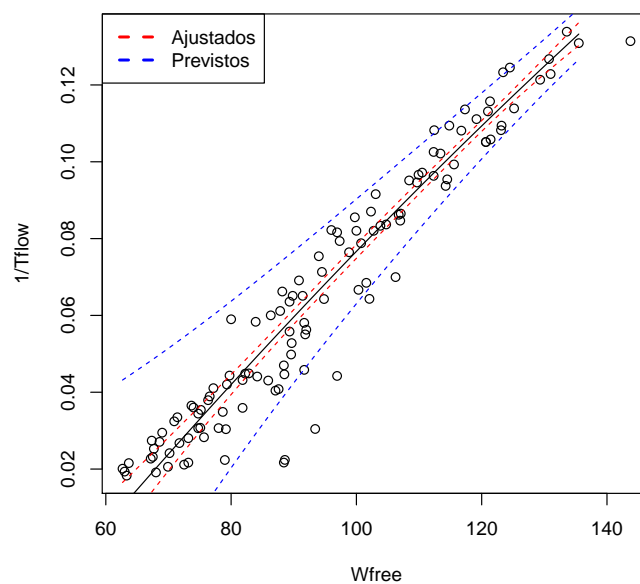
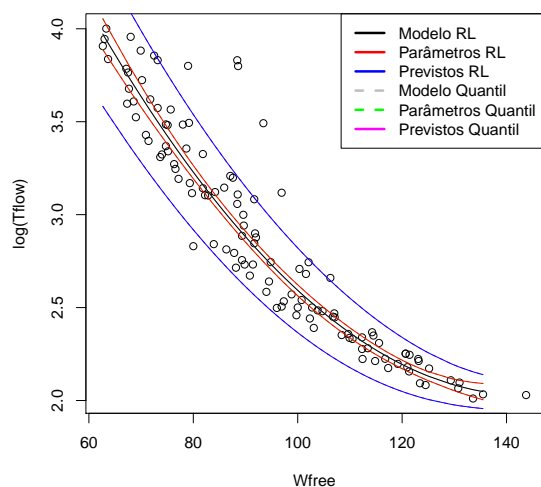
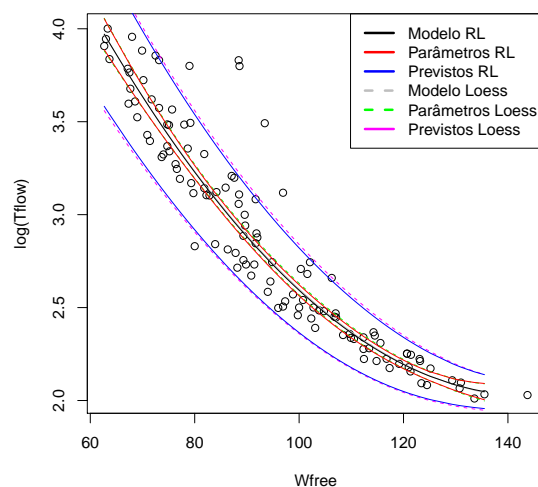


Figura 5.28: Gráfico de dispersão dos dados com a curva de regressão definida pelo modelo (5.17), com os respectivos intervalos de confiança para os valores ajustados da regressão e para os valores previstos.

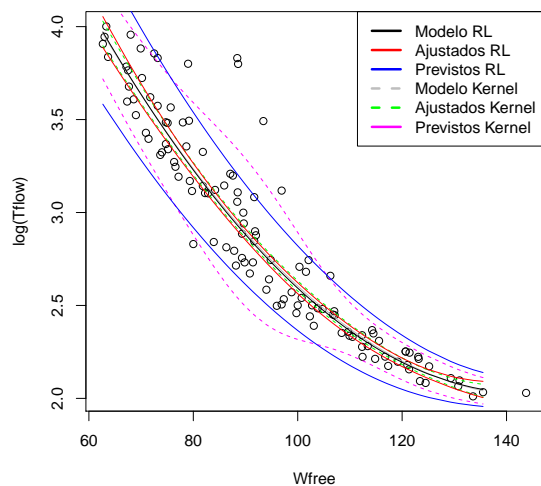


(a) Quantil, Tabela 5.23(2)

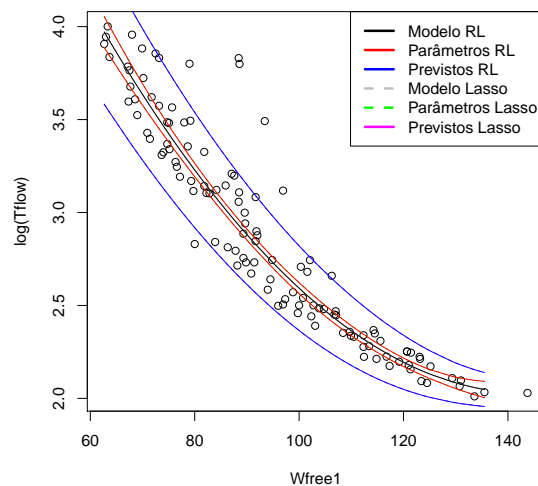


(b) Loess, Tabela 5.23(3)

Figura 5.29: Gráfico de dispersão dos dados com a curva de regressão definida pelo modelo 5.13, com os respectivos intervalos de confiança para os coeficientes de regressão e para os valores previstos em comparação com uma modelação da variância por métodos não paramétricos



(a) Núcleos, Tabela 5.23(4)



(b) Lasso, Tabela 5.23(5)

Figura 5.30: Gráfico de dispersão dos dados com a curva de regressão definida pelo modelo (5.13), com os respectivos intervalos de confiança para os coeficientes de regressão e para os valores previstos em comparação com uma modelação da variância por métodos não paramétricos

Capítulo 6

Conclusões e comentários finais

Neste capítulo são apresentadas as conclusões gerais de todo o trabalho e principais recomendações, reflexo de todos os resultados obtidos experimentalmente, da análise estatística posteriormente realizada e da interpretação que foi possível fazer no contexto de engenharia. As conclusões serão apresentadas focando o objectivo primordial desta dissertação, isto é, a “Modelação de propriedades de pastas de betão autocompactável com comportamento heterocedástico”. Por fim, serão apresentadas algumas propostas consideradas válidas, para que, futuramente se realizem investigações no sentido de aprofundar certos aspectos da vasta área que engloba este tema.

O estudo permitiu concluir que:

- Para valores baixos de água livre temos uma elevada variabilidade do tempo de escoamento e para valores elevados de água livre a variabilidade do tempo de escoamento é baixa;
- para valores de W_{free} superiores a 120 l/m^3 temos que o tempo de escoamento estabiliza, e que portanto amassaduras que tenham um valor superior a este não representam um ganho de eficiência na fluidez da pasta de betão;
- na análise dos modelos estatísticos, o valor de R^2 não é o melhor para aferir do ajustamento deste tipo de modelos e como Willett et. al salientaram (ver [154]), deve-se ter em conta o comportamento dos resíduos e a distribuição destes;
- em termos práticos uma transformação dos dados deve ter em conta o contexto do problema e não deve ser completamente dominada pela optimalidade face a critérios usados na teoria clássica subjacente à modelação (ver [17]);
- a escolha da função que modela a variância não se deve basear no valor de R^2 , mas sim nos resultados obtidos pelo método dos mínimos quadrados ponderados para cada um dos modelos;
- uma má especificação da função que representa a variância provoca problemas com a qualidade dos resíduos como tinha concluído Belsley [18];
- apesar da literatura recomendar a modelação pela via da heterocedasticidade multiplicativa como forma geral de encarar a modelação da variância (ver [95]), neste caso, revelou-se uma opção que não é a mais acertada;

- a influência de outliers deve ser levada em conta na análise dos gráficos de resíduos na tomada de decisões sobre a função que representa a variância [6];
- os gráficos dos resíduos lineares e quadráticos são um bom auxiliar na tomada de decisões sobre a escolha da função que representa a variância [44, 45, 138];
- é necessário ter em atenção se o software usado tem ou não em conta o facto de a transformação do modelo pelo método dos mínimos quadrados ponderado interferir no cálculo do valor de R^2 [154];
- a utilização dos testes sob a heterocedasticidade dos dados deve ser cuidadosa e a informação transmitida por eles deve ser tida em conta para a construção do modelo de variância;
- caso exista informação sobre a forma da heterocedasticidade, a utilização do método dos mínimos quadrados ponderado é preferível aos estimadores de White, pois permite a construção de intervalos de confiança e de previsão mais precisos.

Face ao estudo realizado, apresentamos de seguida recomendações para futuros trabalhos distribuídos por dois tópicos, um deles relativo às opções que foram tomadas na realização de ensaios laboratoriais e um outro relativo à adopção de protocolos experimentais baseados num planeamento estatístico das experiências.

Ensaio de fluidez

A Figura 6.1 mostra-nos a curva de regressão definida pelo modelo (5.13), com os respectivos intervalos de confiança para os coeficientes de regressão e para os valores previstos, com os dados do estudo efectuado em Nunes et. al [110]. Apesar de se ter usado apenas um tipo de cimento no estudo, o modelo obtido para este tipo de cimento seria comum, pelo menos na forma, a outros tipos de cimento. Ora, na verdade, o único tipo de cimento usado nos ensaios foi cimento Portland CEM I 42,5R, enquanto que em Nunes et. al [110] foram contemplados os diversos tipos de cimento existentes para a fabricação de betão auto-compactável, o que se sugere é a realização do mesmo estudo para os restantes tipos de cimento.

Planeamento de experiências

O principal objectivo de um planeamento experimental reside essencialmente na procura orientada de relações entre variáveis e na comparação de níveis de factores, recorrendo ao tratamento estatístico dos dados recolhidos.

Se tivéssemos ao nosso dispor tempo e recursos orçamentais infinitos provavelmente não existiriam grandes necessidade de complicados planeamentos de experiências. Na produção e controle de qualidade queremos controlar o erro e aprender o máximo possível sobre o processo industrial ou a teoria subjacente usando os recursos disponíveis. Numa perspectiva de engenharia, numa fase da experimentação temos de ter em conta que é importante conseguir:

- reduzir o tempo para projectar/desenvolver novos produtos e processos

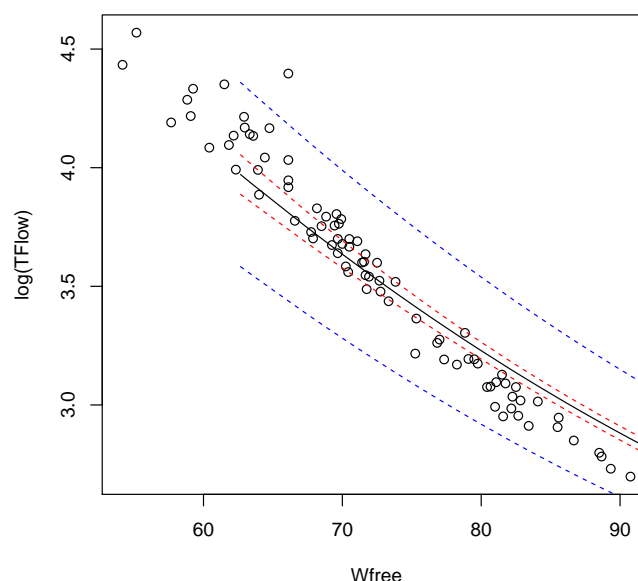


Figura 6.1: Gráfico de dispersão dos dados do estudo efectuado em Nunes et. al [110] e curva de regressão definida pelo modelo (5.13), com os respectivos intervalos de confiança para os coeficientes de regressão e para os valores previstos.

- melhorar o desempenho dos processos existentes
- melhorar a fiabilidade e o desempenho dos produtos
- alcançar robustez do produto e do processo
- realizar a avaliação de materiais, alternativas de projecto, criação de componentes e de sistemas tolerâncias, etc.

Pretende-se sempre afinar ou melhorar o processo. Num mundo globalizado como o de hoje, esta unidade para a competitividade afecta todos nós, tanto como produtores como consumidores.

Existem várias estratégias de experimentação para estudar a influência de tratamentos ou factores sobre determinado sistema. e consequentemente sobre uma dada variável resposta. A mais correcta, nos casos que envolvem vários factores, é realizar uma experiência factorial, na qual se variam simultaneamente todos os níveis dos factores em estudo ao invés de ensaiar um factor de cada vez. Neste caso, estuda-se sucessivamente o efeito sobre a variável resposta de cada um dos factores mantendo os restantes constantes. A maior desvantagem da estratégia "um factor de cada vez", que é muito comum na prática industrial, é não permitir estudar as possíveis interacções entre factores. Por outro lado, a estratégia de "um factor de cada vez" é sempre menos eficiente do que uma experiência factorial.

No trabalho que realizámos para esta Tese foi utilizado um plano factorial 2^{16} para o

planeamento experimental, ao qual o estudo que apresentámos incidiu apenas sobre dois factores, a saber Sp/p e w/c . O plano usado está descrito na literatura para modelos homocedásticos. Trabalhos recentes de Wiens et. al [150] abordam a problemática da definição de planos experimentais para a modelação de relações em presença de heterocedasticidade. Recomendamos um seguimento atento dessa investigação que no futuro poderá dar frutos no melhoramento do planeamento das experiências realizadas.

Apêndice A

Tabela de Registro dos Resultados dos Ensaio

Composição	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
w/c															
Sp/p															
Vw/Vp															
Vs/Vm															
cimento															
Filer calcário															
superplastificante															
água 1															
água 2															
data amassadura															
inicio amassadura															
tfunil 1(sec)															
tfunil 2(sec)															
P1(gr)															
P2(gr)															
P3(gr)															
P4(gr)															
P'1(gr)															
P'2(gr)															
P'3(gr)															
P'4(gr)															
d1 (mm)															
d2 (mm)															
d'1 (mm)															
d'2 (mm)															

Tabela A.1: Tabela de registro dos valores dos ensaios

Bibliografia

- [1] Ali, M. M. and Giaccotto, C. "A study of several new and existing tests for heteroscedasticity in the general linear model", *Journal of Econometrics*, vol.26 (1984), 355-373.
- [2] Agulló, L., Toralles-Carbonari, B., Gettu, R. & Aguado, A., "Fluidity of cement pastes with mineral admixtures and superplasticizer- A study based on the Marsh cone test", *Materials and Structures*, vol.32 (1999), 479-485.
- [3] Aiad, I., "Influence of time addition of superplasticizers on the rheological properties of fresh cement pastes", *Cement and Concrete Research*, vol.33 (2003), 1229 – 1234.
- [4] Aitcin, P. C., *High-performance concrete*, London and New York, E & FN SPON, 1998.
- [5] Akritas, M. G., "Nonparametric estimation in heteroskedastic regression", *Statistics & Probability Letters*, vol.28 (1996), 23-31.
- [6] Alih, E. and Ong, H. C., "An outlier-resistant test for heteroscedascity in linear model", *Journal of Applied Statistics*, (2015), 1-18.
- [7] Amemiya, T., *Advanced Econometrics*, Cambridge, Harvard University Press, 1985.
- [8] Amemiya, T., "Generalized least squares with an estimated autocovariance matrix", *Econometrica*, vol.6 (1973), 723-732.
- [9] Amemiya, T., "A note on a heteroscedastic model", *Journal of Econometrics*, vol.6 (1977), 365-370.
- [10] Asaga, K. & Roy, D. M., "Rheological properties of cement mixes: IV effects of superplasticizers on viscosity and yield stress", *Cement and Concrete Research*, vol.10 (1980), 287-295.
- [11] Aslam, M., "Using heteroscedascity-consistent standard errors for the linear regression model with correlated regressors", *Communications in Statistics-Simulation and Computation*, vol.43 (2013), 2353-2373.
- [12] Aslam, M., Riazb, T. & Altafc, S., "Efficient estimation and robust inference of linear regression models in the presence of heteroscedastic errors and high leverage points", *Communications in Statistics-Simulation and Computation*, vol.43 (2013), 2223–2238.

- [13] Baltagi, B. H., Chang, Y. C., & Li, Q., "Monte Carlo results on several new and existing tests for the error component model", *Journal of Econometrics*, vol.54 (1992), 95-120.
- [14] Baltazar, L. G., Henriques, F. M. A. & Jorne, F., "Optimisation of flow behaviour and stability of superplasticized fresh hydraulic lime grouts through design of experiments", *Construction and Building Materials*, vol.35 (2012), 838–845.
- [15] Bandemer, H., "Methods for checking assumptions in regression models", *Biometrical Journal*, vol.23 (1981), 419–427.
- [16] Bartlett, M. S., "Properties of sufficiency and statistical tests", *Proceedings of the Royal Statistical Society*, vol.160 (1937), 268–282.
- [17] Battese, G. E. & Bonyhady, B. P., "Estimation of household expenditure functions: an application of a class of heteroscedastic regression models", *Economic Record* Vol.57 (1981) 80-85.
- [18] Belsley, D. A. "An investigation of an unbiased correction for heteroskedasticity and the effects of misspecifying the heteroskedastic function", *Journal of Economic Dynamics & Control*, vol.26 (2002), 1379-1396.
- [19] Bickel, P. J., "Using residuals robustly I: tests for heteroscedasticity, nonlinearity", *The Annals of Statistics*, vol.6 (1978), 266-291.
- [20] Bingham, N. H., Fry, J. M. *Regression, Linear Models in Statistics*, Springer-Verlag, 2010.
- [21] Box, G. E. P. & Cox, D. R., "An analysis of transformations", *Journal of the Royal Statistical Society*, 26 (1964), 211–252.
- [22] Box, G. E. P. & Meyer, R. D., "An analysis for unreplicated fractional factorials", *Technometrics*, vol.28 (1986), 11-18.
- [23] Box, G. E. P. & Meyer, R. D., "Dispersion effects from fractional designs", *Technometrics*, vol.28 (1986), 19-27.
- [24] Box, G. E. P., Hunter, W. G. & Hunter JS, *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*, New York, 2nd edition, John Wiley & Sons, 2005.
- [25] Breusch, T. S. and Pagan, A. R. "A simple test for heteroscedasticity and random coefficient variation", *Econometrica* , vol.47 (1977), 1287-1294.
- [26] Brown, M. B., & Forsythe, A. B., "Robust tests for equality of variances", *Journal of the American Statistical Association*, vol.69 (1974), 364–367.
- [27] Brown, J. D., *Linear Models in Matrix Form, A Hands-On Approach for the Behavioral Sciences*, Springer-Verlag, 2014.
- [28] Buse, A., "Goodness of Fit in generalized least squares estimation", *The American Statistician*, vol.27 (1973), 106-119.

- [29] Buse, A., "Goodness-Of-Fit in the seemingly unrelated regressions model-a generalization", *Journal of Econometrics*, vol.10 (1979), 109-113.
- [30] Buse, A., "Tests for Additive heteroskedascity: Goldfeld and Quant revisited", *Empirical Economics*, vol.9 (1984), 199-216.
- [31] Byron, R. P. and Bera A. K. "Least Squares Approximations to Unknown Regression Functions: A Comment", *Biometrika*, vol.70 (1983), 1-10.
- [32] Cai, L., & Hayes, A. F., "A new test of linear hypotheses in ols regression under heteroscedasticity of unknown form", *Journal of Educational and Behavioral Statistics*, vol.33 (2008), 21-40.
- [33] Campos, P. G. S., *Betões Eco-Eficientes com Elevados Teores de Adições Minerais*, Dissertação de Mestrado, Universidade do Minho, Guimarães, 2012.
- [34] Carapeto, M. and Holt, W., "Testing for heteroscedasticity in regression models", *Journal of Applied Statistics*, vol.30 (2003), 13-20.
- [35] Carroll, R. J., "Adapting for heteroscedasticity in linear models", *The Annals of Statistics*, vol.24 (1982), 1224-1233.
- [36] Carroll, R. J. & Welsh, A. H., "A note on asymetry and robustness in linear regression", *International Economic Review*, vol.24 (1983), 255-260.
- [37] Carroll, R. J. & Schneider, H., "A note On Levene's tests for equality of variances", *Statistics & Probability Letters*, vol.3 (1985), 191-194.
- [38] Carroll, R. J. & Cline, D. B. H., "An asymptotic theory for weighted least-squares with weights estimated by replication", *Biometrika*, vol.75 (1988), 35-43.
- [39] Carroll, R. J. & Ruppert, D., *Transformation and Weighting in Regression*, New York, Chapman and Hall, 1988.
- [40] Chatterjee, S. & Hadi, A. S., *Regression, Linear Models in Statistics*, (4.nd edition), John Wiley & Sons, 2006.
- [41] Cheng, Ching-Shui, *Theory of Factorial Design Single- and Multi-Stratum Experiments*, Boca Raton, CRC Press, 2014.
- [42] Chiocchio, G. & Paolini, A. E., "Optimum time for adding superplasticlzers to portland cement pastes", *Cement and Concrete Research*, vol.15 (1985) 901-908.
- [43] Cleveland, W. S., "Robust locally weighted regression and smoothing scatterplots", *Journal of the American Statistical Association* vol.74 (1979), 829-836.
- [44] Cook, R. D., & Weisberg, S. "Diagnostics for heteroscedasticity in regression", *Biometrika*, vol.70 (1983), 1-10.
- [45] Cook, R. D., & Weisberg, S., *Residuals and Influence in Regressions*, New York, Chapman and Hall, 1982.

- [46] Cook, R. D. & Weisberg, S., "Transforming a response variable for linearity", *Biometrika*, vol.81 (1994), 731–737.
- [47] Cooper, R. A. & Weekes, A. J., *Data, Models and Statistical Analysis*, Rowman & Littlefield, 1983.
- [48] Cordeiro, G. M., "Corrected maximum likelihood estimators in linear heteroskedastic regression models", *Brazilian Review of Econometrics*, vol.1 (2008), 1-16.
- [49] Corder, G. W. & Foreman, D. I., *Nonparametric Statistics: A Step-by-Step Approach*, Wiley, 2014.
- [50] Corradini, G., Scoccia, G. & Volpe, R., "Statistical evaluation of mechanical properties of superplasticized concrete" *Cement and Concrete Research*, vol.14 (1984), 375-385.
- [51] Coutinho, A. de S., *Fabrico e propriedades do betão. Vol. I, Laboratório Nacional de Engenharia Civil, Lisboa, 1988.*
- [52] Cribari-Neto, F., "Asymptotic inference under heteroskedasticity of unknown form", *Computacional Statistics and Data Analysis*, vol.45 (2004), 215-233.
- [53] Cribari-Neto, F., Ferrari, S. L. P., & Cordeiro, G. M., "Improved heteroscedasticity-consistent covariance matrix estimators", *Biometrika*, vol.87 (2000), 907-918.
- [54] Cribari-Neto, F., & Zarkos, S. G., "Heteroscedasticity-consistent covariance matrix estimators: White's estimator and the bootstrap", *Journal of Statistical Computation and Simulation*, vol.68 (2001), 391-411.
- [55] Dalgaard, P., *Introductory Statistics with R (2.nd edition)*, Springer, 2008.
- [56] Draper, N. R., "The Box-Wetz criterion versus R^2 ", *Journal of the Royal Statistical Society*, vol.147 (1984), 100-103.
- [57] Dufour, J. M., & Dagenais, M. G., "Durbin-Watson tests for serial correlation in regressions with missing observations", *Journal of Econometrics*, vol.27 (1985), 371-381.
- [58] Dufour, J., Khalaf, L., Bernard, J. & Genest, I., "Simulation based finite sample tests for heteroskedascity and ARCH effects", *Journal of Econometrics*, vol.122 (2004), 317-347.
- [59] Durbin, J. & Watson, G. S., "Testing for serial correlation in least squares regression, I", *Biometrika* vol.37 (1950), 409–428.
- [60] Durbin, J. & Watson, G. S., "Testing for serial correlation in least squares regression, II", *Biometrika* vol.38 (1951), 159–179.
- [61] Durbin, J. & Watson, G. S., "Testing for serial correlation in least squares regression, III", *Biometrika* vol.58 (1971), 159–179.

- [62] Ferrari, L., Kaufman, J., Winnefeld, F. & Plank, J., "Multi-method approach to study influence of superplasticizers on cement suspensions", *Cement and Concrete Research* vol.41 (2011), 1058-1066.
- [63] Forsberg, L., Jonsson, B. & Kristiansson, U., "On type I errors after a preliminary test for heteroscedasticity", *The Statistician*, vol.48 (1999), 63-72.
- [64] Fox, J. & Weisberg, S. *An R Companion to Applied Regression* (2.nd edition), Sage, 2011.
- [65] Geiker, M., "Durability Design of Concrete Structures Minimising Total Life Cycle Costs - Considerations and Examples", *Concrete Durability and Repair Technology: Proceedings of the International Conference Held at the University of Dundee, Scotland*, Edit. R. K. Dhir, M. J. McCarthy, 1999.
- [66] Gentle, J. E., *Matrix Algebra: Theory, Computations, and Applications in Statistics*, Springer-Verlag, 2007.
- [67] Giaccotto, C. and Sharma, S. C. "Jackknife tests for heteroscedasticity in the general linear model", *Australian Journal of Statistics*, vol.30 (1988), 200-216.
- [68] Glejser, H. "A new test for heteroskedacity", *Journal of American Statistical Association*, vol.64 (1969), 316-323.
- [69] Green, W. H., *Econometric Analysis*, (5.nd Edition), New Jersey, Prentice Hall, 2003.
- [70] Godfrey, L. G. "Testing for heteroskedasticity and predictive failure in linear regressions models", *Oxford Bulletin of Economics and Statistics*, vol.70 (2008), 415-429.
- [71] Godfrey, L. G., Orme C. D. & Santos Silva, J. M. C., "Simulation-based tests for heteroskedasticity in linear regression models: Some further results", *Econometrics Journal*, vol.9 (2006), 76-97.
- [72] Golaszewski, J. & Szwabowski, J., "Influence of superplasticizers on rheological behavior of fresh cement mortars", *Cement and Concrete Research*, vol.34 (2004), 235-248.
- [73] Goldfeld, S. M., Quandt, R. E., "Some tests for homoscedasticity", *Journal of the American Statistical Association*, vol.60 (1965), 539-547.
- [74] Gomes, P. C. C., *Optimization And Characterization Of High-Strength Self-Compacting Concrete*, Ph.D. Thesis, Universitat Politècnica Catalunya, Escola Tècnica Superior D'Enginyers De Camins, Barcelona, 2002.
- [75] Hair, J. F., Tatham, R. L., Anderson, R. E. & Black, W., *Multivariate Data Analysis*, (7.nd Edition), New Jersey, Prentice Hall, 2010.
- [76] Han, D. & Ferron, R. D., "Effect of mixing method on microstructure and rheology of cement paste", *Construction and Building Materials* vol.93 (2015), 278-288.
- [77] Hanehara, S., & Yamada, K., "Rheology and early age properties of cement systems", *Cement and Concrete Research*, vol.21 (2008), 175-195.

- [78] Harrell, F. E., *Regression Modeling Strategies With Applications to Linear Models, Logistic Regression and Survival Analysis*, New York, Springer (2001).
- [79] Harrison, M. J., "A simple beta-approximate test for heteroscedasticity based on ordinary least squares residuals", *Irish Journal of Agricultural Economics and Rural Sociology*, Vol.7 (1979), 185-199.
- [80] Harrison, M. J., "A comparison of the bounds, beta-approximate, and exact variants of 2 tests for heteroscedasticity based on ordinary least-squares residuals", *Economic and Social Research Institute, Economic and Social Review*, vol.12 (1981), 235-252.
- [81] Harrison, M. J. & McCabe, B. P. M., "A test for heteroscedasticity based in ordinary least squares residuals", *Journal of the American Statistical Association*, vol.74 (1979), 494-499.
- [82] Harvey, A. & Collier, P., "Testing for functional misspecification in regression analysis", *Journal of Econometrics*, vol.6 (1977), 103-119.
- [83] Harvey, A. C., "Estimating regression models with multiplicative heteroscedasticity", *Econometrica*, vol.44 (1976), 461-465.
- [84] Hinkley, D. V., "Jackknifing in unbalanced situations", *Technometrics*, vol.19 (1977), 285-292.
- [85] Hollander, M., Wolfe, D. A. & Chicken, E., *Nonparametric Statistical Methods*, New York, John Wiley & Sons, (2014).
- [86] Hooper, P. M., "Iterative weighted least squares estimation in heteroscedastic linear models", *Journal of the American Statistical Association* vol.88 (1993), 179-184.
- [87] Huber, P. J., "The behavior of maximum likelihood estimates under nonstandard conditions", *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, (1967), 221-233.
- [88] Imhof, P. J., "Computing the distribution of quadratic forms in normal variables", *Biometrika* vol.48 (1961), 419-426.
- [89] Kauerman, G., & Carroll, R. J., "A note on the efficiency of sandwich covariance matrix estimation", *Journal of the American Statistical Association*, vol.96 (2001), 1387-1396.
- [90] Kiliçkale, F. M., & Dogan, G. G., "Performance of concretes produced with superplasticizer", *Journal of Applied Polymer Science*, vol.103 (2007), 3214-3219.
- [91] Kim, E. H., Morse, A., & Zingales, L., "What has mattered to economics since 1970", *Journal of Economic Perspectives* vol.20 (2006), 189-202.
- [92] Kleiber, C. & Zeileis, A., *Applied Econometrics with R*, Springer-Verlag, 2008.
- [93] Koenker, R., "A note on studentizing a test for heteroscedasticity", *Journal of Econometrics*, vol.17 (1981), 107-112.

- [94] Kmenta, J., *Elements of Econometrics*, New York, MacMillan, 1975.
- [95] Kuk, A. Y. C., "Nonparametrically weighted least squares estimation in heteroscedastic linear regression", *Biometrical Journal*, vol.41 (1999), 401-410.
- [96] Kvalseth, T. O., "Cautionary note about R^2 ", *The American Statistician*, vol.39 (1985), 279-285.
- [97] Judge, G. G., Griffiths W. E., Hill R. C., Lütkepohl H. and Lee T-C., *The Theory and Practice of Econometrics*, (2.nd Edition), New York, John Wiley & Sons, 1985.
- [98] Levene, H., "Robust tests for equality of variances", *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, Stanford University Press, 1960.
- [99] Li, K. & Chan, N. N., "Degeneracy in heteroscedastic regression models", *Journal of Multivariate Analysis*, vol.74 (2000), 282-295.
- [100] Linde, A. & Tutz, G., "On association in regression: the coefficient of determination revisited", *Statistics: A Journal of Theoretical and Applied Statistics*, vol.42 (2008), 1-24.
- [101] Lyon, J. D. & Tsay, C. L., "A comparison of tests for heteroscedasticity", *Journal of the Royal Statistical Society*, vol.45 (1996), 337-349.
- [102] MacKinnon, J. G. and White H. "Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties", *Journal of Econometrics*, vol.29 (1985), 305-325.
- [103] Maekawa, K., "Comparing the Wald, LR And LM tests for heteroscedasticity in a linear regression model", *Economics Letters*, vol.26 (1988), 37-41.
- [104] Manning, W. G. and Mullahy, J., "Estimating log models: to transform or not to transform", *Journal of Health Economics*, vol.20 (2001), 461-494.
- [105] Mosteller, F. & Tuckey, J. W., *Data Analysis and Regression: A Second Course in Statistics*, Reading, Addison-Wesley, 1977.
- [106] Murteira, B., Ribeiro, C. S., Andrade e Silva, J. & Pimenta, C., *Introdução à Estatística*, Lisboa; Escolar Editora, 2010.
- [107] Murteira, J. M R., Ramalho, E. A., & Ramalho, J. J. S., "Heteroskedasticity testing through a comparion of Wald statistics", *Portuguese Economic Journal*, vol.12 (2013), 131-160.
- [108] Myers R. H., Montgomery C. D., Vining G. G. & Robinson T.J. *Generalized Linear Models with Applications in Engineering and the Sciences* (2.nd edition), John Wiley & Sons, 2010.
- [109] Nguyen, V. H., Rémond, S., Gallias, J. L., Bigas, J. P. & Muller, P., "Flow of Herschel-Bulkley fluids through the Marsh cone", *Journal of Non-Newtonian Fluid Mechanics*, vol.139 (2006), 128-134.

- [110] Nunes, S., Oliveira, P. M., Coutinho, J. S. & Figueiras, J., "Rheological characterization of SCC mortars and pastes with changes induced by cement delevery", *Cement & Concrete Composites*, vol.33 (2011), 103-115.
- [111] Nunes, S. C. B., *Betão auto-compactável: Tecnologia e propriedades*, Tese de Mestrado, Faculdade de Engenharia da Universidade do Porto, Porto, 2001.
- [112] Önder, A. Ö. & Zaman, A. "Robust tests for normality of errors in regression models", *Economic Letters*, vol.86 (2005), 63-68.
- [113] Osbourne, J. W., & Waters, E., "Four assumptions of multiple regression that researchers should always test", *Pratical Assessment, Research, and Evaluation*, vol.8, 1-5.
- [114] Park, R. E., "Estimation of Heteroscedastic Error Terms", *Econometrica* , vol.34 (1966), 888.
- [115] Peiris, M. S., "Efficient estimation of regression models with heteroscedastic errors", *Mathematical Scientist*, vol.38 (2013), 124-128.
- [116] Ramsey, J. B., "Tests for specification errors in classical linear least-squares regression analisys", *Journal of the Royal Statistical Society. Series B*, vol.31 (1969), 350-371.
- [117] Randolph, W. C., "A transformation for heteroscedastic error components regression models", *Economic Letters*, vol.27 (1988), 349-354.
- [118] Rao, C. R., "Estimation of variance and covariance Components - MINQUE theory", *Journal of Multivariate Analysis*, vol.1 (1971), 257-275.
- [119] Rilstone, P., "Some Monte Carlo evidence on the relative efficiency of parametric and semiparametric egls estimators", *Journal of Business & Economic Statistics*, vol.9 (1991), 179-187.
- [120] Roy, RK., *Design of experiments using Taguchi approach: 16 steps to product and process improvement*, New York, John Wiley & Sons, 2001.
- [121] Roy, R. L. & Roussel, N., "The Marsh cone as a viscometer: theoretical analysis and pratical limits", *Materials and Structures*, vol.38 (2005), 25-30.
- [122] Roussel, N. & Roy, R. L., "The Marsh cone: a test or a rheological apparatus?", *Cement and Concrete Research*, vol.35 (2005), 823-830.
- [123] Rubinstein, R. Y., *Simulation and the Monte Carlo Method*, New York, John Wiley & Sons, 1981.
- [124] Samworth, R. J., "Optimal weighted nearest neighbour classifiers", *Annals of Statistics*, vol.40 (2012), 2733-2763.
- [125] Savin, N. E., & White, K. J., "The Durbin-Watson test for serial correlation with extreme sample sizes or many regressors", *Econometrica*, vol.45 (1977), 1989-1996.

- [126] Shen, S. & Mei, C., "Estimation of the variance function in heteroscedastic linear regression models", *Communications in Statistics - Theory and Methods*, vol.38 (2009), 1098-1112.
- [127] Spearman, C., "The proof and measurement of association between two things", *The American Journal of Psychology*, vol.15 (1904), 72-101.
- [128] Smyth, G. K., "An efficient algorithm for REML in heteroscedastic regression", *Journal Graphical and Computacional Statistics*, vol.11 (2002), 836-847.
- [129] Sundberg, R., "The Precision of the Estimated Generalized Least. Squares Estimator in Multivariate Calibration", *Scandinavian Journal Of Statistics*, vol.23 (1996), 257—274.
- [130] Surekha, K. & Griffiths, W. E., "A monte carlo comparison of some bayesian and sampling theory estimators in two heteroscedastic error models", *Communications in Statistics - Simulation and Computation*, vol.13 (1984), 85-105.
- [131] Szroeter, J., "A class of parametric tests for heteroscedasticity in linear econometric models", *Econometrica*. vol.46 (1978), 1311-1327.
- [132] Theil, H., *Principles of Econometrics*, (1971), Wiley, New York.
- [133] Theil, H., "Estimates and their sampling variance of parameters of certain heteroscedastic distributions", *Review of the International Statistical Institute*, vol.19 (1951), 141-147.
- [134] Thursby, J. G., "Misspecification, heteroscedascity, and the Chow and Goldfeld-Quandt Tests", *The Review of Economics and Statistics*, vol.64 (1982), 314-321.
- [135] Thursby, J. G., & Schmidt, P., "Some properties of tests for specification error in a linear regression model", *Journal of the American Statistical Association*, vol.72 (1977), 635-641.
- [136] Tibshirani, R., "Regression shrinkage and selection via the LASSO", *Journal of the Royal Statistical Society*, vol.58 (1996), 267-288.
- [137] Tuckey, J. W., , *Exploratory Data Analysis*, Reading, Addison-Wesley, 1977.
- [138] Tsai, C. L., Cai, Z. & Wu X., "The examination of residual plots", *Statistica Sinica*, vol.8 (1998), 445-465.
- [139] Tse, Y. K., "Testing for linear and log-linear regressions with heteroscedacity", *Economic Letters*, vol.16 (1984), 63-69.
- [140] Tsurumi, H. & Sheflin, N., "Some tests for the constancy of regressions under heteroscedacity", *Journal of Econometrics*, vol.27 (1985), 221-234.
- [141] Utts, J.M., "The Rainbow test for lack of fit in regression", *Communications in Statistics - Theory and Methods*, vol.11 (1982), 1801-1815.

- [142] Venables, W. & Ripley, B. , *Modern Applied Statistics with S*, (4.nd edition), Springer-Verlag, 2002.
- [143] Verbyla, A. P. "Modelling variance heterogeneity: residual maximum likelihood and diagnostics", *Journal of the Royal Statistical Society*, vol.55 (1993), 493-508.
- [144] Víšek, J. Á., "Considtency of the least weighted squares under heteroscedasticity", *Kybernetika*, vol.47 (2011), 179-206.
- [145] Víšek, J. Á., "Empirical distribution function under heteroscedasticity", *Statistics: A journal of Theoretical and Applied Statistics*, vol.45 (2011), 497-508.
- [146] Waldman, D. M. "A note on algebric equivalence of White's test and a variation of the Godfrey/Breuch-Pagan test for heteroscedasticity", *Economics Letters*, vol.13 (1983), 197-200.
- [147] Weisberg, S. *Applied linear regression* (3.nd edition), New Jersey, John Wiley & Sons, 2011.
- [148] Wetzel, N. R., "Graphical data modeling methods using ceres plots", *Journal of Statistical Computation and Simulation*, vol.54 (1996), 1-3.
- [149] Wiens, D. P., "Minimax robust designs and weights for approximately specified regression models with heteroscedastic errors", *Journal of the American Statistical Association*, vol.93 (1998), 1440-1450.
- [150] Wiens, D. P. & Li, P., "V-optimal design for heteroscedastic regression", *Journal of Statistical Planning and Inference*, vol.145 (2014), 125-138.
- [151] Wilcox, R. R., "Testing hypotheses about regression parameters when the error term is heteroscedastic", *Biometrical Journal*, vol.41 (1999), 411-426.
- [152] Wilcox, R. & Keselman H., "Modern regression methods that can substantially increase power and provide a more accurate understanding of associations", *European Journal of Personality*, vol.32 (2012), 165-174.
- [153] White, H. "A heteroskedascity-consistent covariance matrix and a direct test for heteroskedascity", *Econometrica*, vol.48 (1980), 817-838.
- [154] Willett, J. B. & Singer, J. D., "Another cautionary note about R²: Its Use in Weighted Least-Squares Regression Analysis", *The American Statistician*, vol.42 (1988), 236-238.
- [155] Yamada,, K., "Basics of analytical methods used for the investigation of interaction mechanism between cements and superplasticizers", *Cement and Concrete Research* vol.41 (2011), 793-798.
- [156] Yin, Y. & Carrol, R. J., "A diagnostic for heteroscedacity based on the spearman rank correlation", *Statistics & Probability Letters*, vol.10 (1990), 69-76.